UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS
Departamento de Bioquímica y Biología Molecular I

TESIS DOCTORAL

Avances en el estudio de la enzimas implicadas en la ruta de degradación del anillo esteroideo en *Rodococcus ruber Chol-4*

New insights into the enzymes involved in the steroid ring degradation pathway of *Rodococcus ruber Chol-4*

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Flor Govinda Guevara Acosta

Directora

Juana María Navarro Llorens

Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS QUÍMICAS

Departamento de Bioquímica y Biología Molecular I

TESIS DOCTORAL

Avances en el estudio de las enzimas implicadas en la ruta de degradación del anillo esteroideo en *Rhodococcus ruber* Chol-4

New insights into the enzymes involved in the steroid ring degradation pathway of *Rhodococcus ruber* Chol-4

FLOR GOVINDA GUEVARA ACOSTA

Directora:

DRA. JUANA MARÍA NAVARRO LLORENS

Madrid, 2017

Diseño de la Portada y Contraportada por

Elena Vargas

El trabajo descrito en esta tesis ha sido llevado a cabo en el Departamento de Bioquímica y Biología Molecular I de la Facultad de Ciencias Biológicas de la Universidad Complutense de Madrid. La investigación ha sido financiada por el proyecto BFU2009-11545-C03-02, BIO2012-39695-CO2-01 y RTC-2014-2249-1 del Ministerio de Educación y Ciencia y MINECO.

A mi abuelita

A mi padre

A Rafa

AGRADECIMIENTOS

Me gustaría expresar mi más profundo agradecimiento a todas las personas que han colaborado de manera directa o indirectamente a la realización de esta Tesis Doctoral.

Ante todo, me gustaría agradecer la guía, seguimiento y apoyo recibido en todo momento por mi directora de tesis la Dra. Juana María Navarro Llorens. También me gustaría agradecer especialmente al Prof. Julián Perera por el apoyo, los ánimos y toda la ayuda recibida a lo largo de la elaboración de este proyecto y a cada una de las personas que forman parte de su grupo de investigación. A Amaya por su apoyo incondicional. A todas las personas del Departamento de Bioquímica y Biología Molecular I por haberme dado la oportunidad de integrarme en este centro (al Director del Departamento Jesús Pérez-Gil, a la coordinadora de Doctorado Maria Antonia Lizarbe y un largo etc.) y por compartir con ellos ciencia. Especialmente, al grupo de la Doctora Isabel de la Mata y el Doctor Miguel Arroyo que han sido testigos de todos mis avatares.

Agradezco de igual manera al doctor Jose Luis García y a los miembros de su grupo del CIB (Beatriz, Carmen, Julia y demás) por su apoyo tan necesario a lo largo de mi periodo predoctoral, al grupo del Catedrático Jose María Luengo de la Universidad de León y a la empresa GADEA por su contribución al proyecto que tenemos entre manos y que ha financiado mi formación. De igual manera agradezco a los servicios de secuenciación Secugen, al CAI de genómica de Biológicas y a Rodrigo Velasco especialmente que casi funciona como un CAI-para-todo.

Finalmente, un agradecimiento profundo al cariño, comprensión y ánimo recibidos de mi familia y amigos.

A todos ellos, muchas gracias.

ÍNDICE

ABBREVIATIONS

µg	microgram
µL	microlitre
µM	micromolar
4-BNC	4-pregnen-3-one-20b-carboxylic acid
5β-AD	5β-androstane-3,17-dione
5α-Tes	17β-hydroxy-5α-androstan-3-one
9OH-AD	9α-hydroxy-4-androstene-3,17-dione
9OH-ADD	9α-hydroxy-1,4-androstadiene-3,17-dione
AD	4-androstene-3,17-dione
ADD	1,4-androstadiene-3,17-dione
BLAST	basic local alignment search tool
CHO	cholest-5-en-3β-ol (cholesterol)
CoA	Coenzyme A
CTAB	cetyltrimethyl ammonium bromide
DCPIP	2,6-dichlorophenol-indophenol
DHEA	dehydroepiandrosterone
DO600nm	optical density at 600nm
DOC	deoxycorticosterone
FAD	flavin adenine dinucleotide
HPLC	high-performance liquid chromatography
HSA	3-hydroxy-9,10-secoandrost-1,3,5(10)-triene-9,17-dione
kDa	1000 Dalton
KshAB	3-ketosteroid-9α-hydroxylase
KstD	3-ketosteroid-Δ^1-dehydrogenase
LB	Luria-Bertani
Mb	mega base pair
mg	miligram
mL	mililitre
mM	milimolar
MSC	multiple cloning site
NAD	nicotinamide adenine dinucleotide
NCBI	The National Center for Biotechnology Information
nm	nanometer

ORF	open reading frame
PCA	protocatechuic acid
PDB	Protein Data Banck
PFGE	pulsed field gel electrophoresis
RAST	rapid annotation using subsystem technology
rpm	revolutions per minute
RT-PCR	reverse transcription polymerase chain reaction
RT-qPCR	reverse transcription polymerase chain reaction
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TCA	tricarboxylic acid cycle
TSS	transcriptional start site
Δ1,4-BNC	3-oxo-23,24-bisnorchola-1,4-dien-22-oic acid
Ω	ohmio

Aminoacid code

Three letter	One leter	Name
Ala	A	Alanin
Arg	R	Arginin
Asn	N	Asparagin
Asp	D	Aspartic acid
Cys	C	Cystein
Gln	Q	Glutamine
Glu	E	Glutamic acid
Gly	G	Glicyne
His	H	Hystidine
Ile	I	Isoleucine
Leu	L	Leucine
Lys	K	Lysine
Met	M	Methionine
Phe	F	Phenylalanine
Pro	P	Proline
Ser	S	Serine
Thr	T	Threonine
Trp	W	Tryptophan
Tyr	Y	Tyrosin
Val	V	Valine

I. SUMMARY

1. Introduction

Rhodococcus ruber strain Chol-4 is an actinobacteria isolated from a sewage sludge sample. It is able to grow in minimal medium supplemented with steroids and aromatic compounds, showing a large catabolic capacity. The degradation of aromatic compounds and steroids helps to maintain the global carbon cycle, and it also has an increasing number of biotechnological applications from biodegradation of pollutants to production of pharmaceutical compounds. The genus rhodococci contains multiple homologues of catabolic genes, which may be the reason of their wide and versatile catabolic capabilities. However, the redundancy of genes in the actinobacterial genomes results in a serious complication for metabolic engineering to obtain intermediates of industrial interest.

The 3-ketosteroid-Δ^1-deydrogenase activity (KstD) is a key enzyme in the general scheme of the bacterial steroid catabolism in combination with the 3-Ketosteroid-9α-Hydroxylase (KshAB), being both responsible of the steroid nucleus (rings A/B) breakage. KshAB initiates the opening of the steroid ring by the 9α-hydroxylation of the C9 carbon of 4-ene-3-oxosteroids (e.g. AD) or 1,4-diene-3-oxosteroids (e.g. ADD), transforming them into 9α-hydroxy-4-androsten-3,17-dione (9OHAD) or 9α-hydroxy-1,4-androstadiene-3,17-dione (9OHADD) respectively. KstD converts 4-ene-3-oxosteroids (e.g. AD) or 9-hydroxy-4-ene-3-oxosteroids (9OH-AD) to 1,4-diene-3-oxosteroids (e.g. ADD) or 9-hydroxy-1,4-diene-3-oxosteroids (e.g. 9OH-ADD) by trans-axial elimination of the C-1(α) and C-2(β) hydrogen atoms.

2. Objectives and Results

The aim of this work is the functional characterization of the catabolic enzymes in the *Rhodococcus* steroid ring A/B degradation prior to the developing of different biotechnological applications.

First at all, the genome of Chol-4 was explored and annotated in order to identify several gene clusters involved in its catabolic skills. The studies have revealed the presence of seven out of the eight clusters codifying for central aromatic degradation pathways previously described in *R. jostii* RHA1. Several peripheral pathways have been also identified indicating its high biodegradation/biotransformation potential. Mutants of two predicted clusters (*pac* and *nar* genes) were made to confirm their identification: these mutants lost the ability to grow in presence of its corresponding substrate (protocatechuate and naphthalene) as the only source of energy and carbon.

The redundancy of enzymes, related to the steroid degradation, found in the actinobacterial genomes results in a serious difficulty for the metabolic engineering of this catabolic pathway to obtain intermediates of industrial interest. For instance, at least three putative 3-ketosteroid Δ^1-dehydrogenase ORFs (*kstD1*, *kstD2* and *kstD3*) have been identified and functionally confirmed in *Rhodococcus ruber* strain Chol-4. KstD1 and KstD2 display a wide range of substrate preferences regarding to well-known intermediaries of the cholesterol degradation pathway (9OH-AD and AD) and other steroid compounds. KstD3 shows a narrower substrate range with a preference for saturated substrates. KstDs differences in their catalytic properties could be somehow related to structural differences revealed by a preliminary structural modelling. The expression of *kstD* genes was followed by RT-PCR and RT-qPCR. All the *kstD* genes were transcribed under all the conditions assayed, although an additional induction could be observed for *kstD1* in AD and *kstD3* in cholesterol.

Three homologous *kshA* and one *kshB* gene were also characterized, each one in different genomic regions of *R. ruber* strain Chol-4. A set of data has been presented, that helps to understand their specific roles in this strain, including: i) description of the KshAB enzymes ii) construction and characterization of *ΔkshB* and single, double and triple *ΔkshA* mutants in *R. ruber* iii) growth studies of the above mutant strains on different substrates and iv) their genetic complementation and biotransformation assays. Our results showed that KshA2 isoform is needed for the degradation of steroid substrates with short side chain, while KshA3 works on those molecules with longer side chains. KshA1 is a more versatile enzyme related to the cholic acid catabolism, although it also collaborates with KshA2 or KshA3 activities in the catabolism of steroids.

In the Doctoral work here presented, some biotechnological applications have also been tried. Concretely, intensifying the KstD strength in a *Rhodococcus* strain in order to obtain a dehydrogenation-enhanced strain.

3. Conclusions

Briefly, results presented in this work show that *R. ruber* strain Chol-4 has a high biotechnological interest due to its metabolic potential (it is able to grow in minimal medium with steroids e.g. cholesterol, phytosterols, DHEA, bile acids such as cholic acid or aromatic compounds e.g. benzoate, naphtalene, gentisate as the only source of carbon and energy). Our genome analysis provides a valuable guide for future in depth molecular or physiological studies for *R. ruber* microbiology and provide a basis for better exploiting this strain for biotechnology.

This study also has provided biochemical and genetic insights into the three KstD and KshA proteins found in *R. ruber*. Although the profile of several substrates of KstD1 and KstD2 was quite similar *in vitro*, the role *in vivo* for them is different: KstD2 is essential for the growth on testosterone, cholic acid and AD while KstD1 has not a defined role in this strain, although its expression was induced when growing on AD. KstD3 that showed a higher preference for saturated steroid substrates is involved in the cholesterol and phytosterols growth in this strain.

On the other hand, our results point to a role of KshA2 isoform in the degradation of steroid substrates such as AD, ADD and testosterone. KshA3 is mainly involved in the catabolism of steroids with longer side chains such as cholesterol and lastly KshA1 would collaborate with other KshA activities in the catabolism of other steroids (except testosterone).

Moreover, our results suggest that the side chain degradation would be KshAB-independent in this strain.

Some biotechnological applications of these findings are under work.

II. RESUMEN

RESUMEN

1. Introducción

Rhodococcus ruber Chol-4 es una actinobacteria aislada a partir de una muestra de lodos de depuradora. Esta cepa es capaz de crecer en medio mínimo suplementado con esteroides y compuestos aromáticos, mostrando una gran capacidad catabólica. La degradación de compuestos aromáticos y esteroides ayuda a mantener el ciclo global del carbono, además de tener un número creciente de aplicaciones biotecnológicas que van desde la biodegradación de contaminantes hasta la producción de compuestos farmacéuticos. El género rhodococci contiene múltiples genes homólogos, lo que le confiere una amplia y versátil capacidad catabólica. Sin embargo, la redundancia de genes en los genomas actinobacterianos resulta una seria complicación para obtener intermediarios de interés industrial mediante ingeniería metabólica.

La actividad de la enzima 3-cetosteroide-Δ^1-deshidrogenasa (KstD) en combinación con la 3-cetosteroide-9α-hidroxilasa (KshAB) es clave en el esquema general del catabolismo bacteriano de esteroides, ambas son responsables de la rotura del núcleo esteroide A/B. KshAB inicia la apertura del anillo de esteroides mediante la 9α-hidroxilación del carbono C9 de 4-eno-3-oxosteroides (p. ej. AD) o 1,4-dieno-3-oxosteroides (p. ej. ADD), transformándolos en 9α-hidroxi-4-androsten-3,17-diona (9OHAD) o 9α-hidroxi-1,4-androstadien-3,17-diona (9OHADD), respectivamente. KstD convierte 4-eno-3-oxoesteroides (p. ej. AD) o 9-hidroxi-4-eno-3-oxoesteroides (p. ej. 9OH-AD) a 1,4-dieno-3-oxoesteroides (p. ej. ADD) o 9-hidroxi-1,4-dieno-3-oxoesteroides (p. ej. 9OH-ADD) por eliminación transaxial de los átomos de hidrógeno C-1 (α) y C-2 (β).

2. Objetivos y resultados

El objetivo de este trabajo es la caracterización funcional de las enzimas catabólicas de degradación del anillo esteroideo A/B en *Rhodococcus ruber* Chol-4 para el desarrollo posterior de aplicaciones biotecnológicas.

En primer lugar, se analizó y anotó el genoma de Chol-4, con el fin de identificar *clusters* génicos implicados en sus habilidades catabólicas. Los estudios han revelado la presencia de siete de los ocho *clusters* que codifican las vías centrales de degradación de compuestos aromáticos, descritas anteriormente en *R. jostii* RHA1. También se han identificado varias vías periféricas de estos compuestos que indican su alto potencial de biodegradación/biotransformación. Se hicieron mutantes de dos de los *clusters* predichos (genes *pac* y *nar*) para confirmar su

identificación. Estos mutantes perdieron la capacidad de crecer en presencia de su sustrato correspondiente (protocatecuato y naftaleno) como única fuente de energía y carbono.

La redundancia de enzimas relacionadas con la degradación de esteroides que se ha encontrado en los genomas actinobacterianos, da como resultado una seria dificultad en la implementación de ingeniería metabólica para obtención de intermediarios de interés industrial. Por ejemplo, se han identificado y confirmado funcionalmente al menos tres ORF putativos de 3-cetosteroides-Δ^1-deshidrogenasa (*kstD1, kstD2* y *kstD3*) en la cepa de *R. ruber* Chol-4. KstD1 y KstD2 muestran una amplia gama de preferencias de sustrato con respecto intermediarios conocidos de la ruta de degradación del colesterol (p. ej. 9OH-AD y AD) y además de otros compuestos esteroideos. La KstD3 muestra un rango de sustrato más estrecho con una preferencia por sustratos saturados. Las diferencias de las propiedades catalíticas de cada KstD podrían estar de alguna manera relacionadas con las diferencias estructurales encontradas en un proceso de modelización estructural preliminar. La expresión de los genes *kstD* fue seguida por RT-PCR y RT-qPCR. Todos los genes kstD se transcribieron en todas las condiciones ensayadas, aunque una inducción adicional se observó para kstD1 en AD y kstD3 en colesterol.

También en este estudio se han caracterizado tres genes homólogos *kshA* y un gen *kshB*, cada uno de ellos en diferentes regiones genómicas de *R. ruber* Chol-4. En este trabajo se recoge un conjunto de datos que ayudan a comprender sus funciones específicas en esta cepa, incluyendo: i) descripción de las enzimas KshAB ii) construcción y caracterización de mutantes ΔkshB y ΔkshA simples, dobles y triples en *R. ruber* iii) estudios de crecimiento de las cepas mutantes en diferentes sustratos y iv) ensayos de complementación genética y biotransformación de los mutantes. Nuestros resultados mostraron que la isoforma KshA2 es necesaria para la degradación de los sustratos esteroides con cadena lateral corta, mientras que KshA3 trabaja en las moléculas con cadenas laterales más largas. KshA1 es una enzima más versátil relacionada con el catabolismo ácido cólico, aunque también colabora con actividades KshA2 o KshA3 en el catabolismo de esteroides.

En el trabajo de Doctorado aquí presentado, también se han probado algunas aplicaciones biotecnológicas. Concretamente, intensificando la actividad KstD en una cepa de *Rhodococcus* para obtener una cepa con la función de deshidrogenación mejorada.

3. Conclusiones

En resumen, los resultados presentados en este trabajo muestran que la cepa de *R. ruber* Chol-4 presenta un alto interés biotecnológico debido a su potencial metabólico. Es capaz de crecer

en medio mínimo con esteroides, por ejemplo colesterol, fitoesteroles, DHEA, ácidos biliares tales como ácido cólico o compuestos aromáticos, por ejemplo, benzoato, naftaleno, gentisato como única fuente de carbono y energía. Nuestro análisis del genoma proporciona una valiosa guía para futuros estudios moleculares o fisiológicos en profundidad para la microbiología de *R. ruber* y proporciona una base para una mejor explotación biotecnológica de esta cepa.

Este estudio también ha proporcionado información bioquímica y genética de las tres proteínas KstD y KshA encontradas en *R. ruber*. Aunque el perfil de varios sustratos de KstD1 y KstD2 fue bastante similar *in vitro*, el papel *in vivo* para cada una es diferente: KstD2 es esencial para el crecimiento en testosterona, ácido cólico y AD; mientras que KstD1 no tiene un papel definido en esta cepa, aunque su expresión se indujo al crecer en AD; KstD3 mostró una mayor preferencia por los sustratos de esteroides saturados y está involucrada en el crecimiento de colesterol y fitoesteroles.

Por otro lado, nuestros resultados apuntan a que la isoforma KshA2 interviene en la degradación de los sustratos esteroideos como AD, ADD y testosterona. KshA3 está principalmente involucrada en el catabolismo de los esteroides con cadenas laterales más largas como el colesterol y por último KshA1 colaboraría con otras actividades KshA en el catabolismo de otros esteroides (excepto la testosterona).

Además, nuestros resultados sugieren que la degradación de la cadena lateral sería independiente de KshAB en esta cepa.

Algunas aplicaciones biotecnológicas de estos hallazgos se están desarrollando en la actualidad.

III. INTRODUCCIÓN

INTRODUCCION

1. Actinomicetos

Los actinomicetos son bacterias heterótrofas, Gram-positivas, filamentosas y morfológicamente diversas que habitan en una amplia variedad de ecosistemas, tanto marinos como terrestres (Sharma 2014). Pertenecen al orden Actinomicetales, filo Actinobacteria, una de las unidades taxonómicas más grandes de entre los 18 linajes principales del Dominio Bacteria (Ventura, Canchaya *et al.* 2007). Se caracteriza por tener un alto contenido en guanina (G) y citosina (C) en su DNA (Ventura, Canchaya *et al.* 2007). A este orden pertenecen una gran variedad de especies entre las cuales encontramos patógenos (p. ej. *Mycobacterium tuberculosis*), cepas que son capaces de producir compuestos bioactivos (p. ej. *Streptomyces*, *Amycolatopsis*) e incluso algunas morfológicamente se asemejan a hongos por su formación de micelio e hifas (p. ej. *Streptomyces spp*.).

Los actinomicetos se identifican y clasifican en función de su análisis molecular, quimitaxonómico y fenotípico. Se dividen en órdenes, subórdenes, familias y géneros. Estos dos últimos grupos taxonómicos están en constante cambio debido al creciente número de cepas que se descubren e identifican (Goodfellow, Alderson *et al.* 1998, Koerner, Goodfellow *et al.* 2009). Son bacterias aerobias o microaerófilas (Embley y Stackebrandt 1994) y sus membranas celulares pueden contener ácidos micólicos. A los organismos de este filo que contienen ácidos micólicos se les denomina “micolata” (Gebhardt, Meniche *et al.* 2007) y pertenecen al suborden Corynebacterineae. Los ácidos micólicos son ácidos grasos β-hidroxi de alto peso molecular, que presentan α-ramificaciones en dos posiciones con cadenas alquilo largas (Koerner, Goodfellow *et al.* 2009). A los ácidos micólicos se les atribuye varias funciones, entre las que se pueden mencionar el facilitar la toma de sustratos hidrófobos a la célula (Martínková, Uhnakova *et al.* 2009) y contribuir a las propiedades de adhesión de bacterias (Sokolovská, Rozenberg *et al.* 2003). También son responsables de la formación de una barrera de permeabilidad que evita la entrada de antibióticos y otras moléculas en la célula.

Dentro del orden Actinomicetales, suborden Corynebacterineae, encontramos la familia Nocardiaceae a la que pertenece el género *Rhodococcus* (Zhi, Li *et al.* 2009).

2. El género Rhodococcus

El género *Rhodococcus* fue propuesto inicialmente por Zopf en 1891 para bacterias que producen un pigmento rojo y desde entonces la sistemática del género ha variado mucho

(Goodfellow, Alderson *et al.* 1998). Los rhodococci son organismos aerobios, quimótrofos, no móviles, no esporulantes, con un contenido de G + C en su DNA de 63 a 73% y algunos presentan capacidad para fijar nitrógeno y formación de biofilms (Goodfellow, Alderson *et al.* 1998, Kuyukina y Ivshina 2010). Las especies conocidas del género *Rhodococcus* (Figura 1) se han aislado principalmente del suelo, pero también son comunes en otros entornos como: el agua, las espumas de lodos activados, las tripas de los artrópodos hematófagos, el estiércol de herbívoros (Goodfellow 1989), en ciertos animales, rocas (Bell, Philp *et al.* 1998) y desiertos (Hernandez, Mohn *et al.* 2008).

Los rhodococci pueden ser simbiontes (p. ej. *Rhodococcus rhodnii*), patógenos de plantas (p. ej. *Rhodococcus fascians*) y patógenos de animales, incluyendo seres humanos (p. ej. *Rhodococcus equi*) (Bell, Philp *et al.* 1998). Tienen forma bacilar o cocoidal en el inicio de su ciclo de vida (Goodfellow, Alderson *et al.* 1998). A medida que se desarrollan pueden mostrar una gran heterogeneidad morfológica que depende de la cepa, de las condiciones de crecimiento y/o de la fase de crecimiento (Larkin, De Mot *et al.* 1998). Los bacilos y cocos pueden germinar y dar lugar a filamentos, filamentos ramificados, sinemas aéreas, hifas aéreas de tamaño microscópico o hifas ramificadas. La fragmentación de estas estructuras da inicio a una nueva generación (Goodfellow, Alderson *et al.* 1998). Curiosamente, en varias cepas de *Rhodococcus* se ha evidenciado la formación de arquitecturas aéreas especializadas -similares a hongos diminutos- desarrolladas para la utilización de alquil fenoles vaporizados (p. ej. *p*-cresol) a modo de fuente de carbono (Veeranagouda, Lim *et al.* 2009).

Figura 1. Fotografía de una placa de LB Agar con *Rhodococcus ruber*

Rhodococcus posee un metabolismo oxidativo con capacidad de utilizar una amplia variedad de compuestos orgánicos como única fuente de carbono y energía, incluyendo hidrocarburos gaseosos (Bell, Philp *et al.* 1998), xenobióticos (compuestos cuya estructura química en la naturaleza es poco frecuente o inexistente) tales como hidrocarburos alifáticos, aromáticos y

clorados; compuestos oxigenados, nitroaromáticos, nitrilo, halogenados y heterocíclicos (Martínková, Uhnakova *et al.* 2009).

Adicionalmente, los miembros del género *Rhodococcus* son capaces de degradar gran variedad de compuestos contaminantes, tóxicos y recalcitrantes tales como los bifenilos policlorados, muchos de ellos presentes en fluidos aislantes, pinturas y plásticos (Martínková, Uhnakova *et al.* 2009); explosivos como la hexahidro-1,3,5-trinitro-1,3,5-triazina (Seth-Smith, Edwards *et al.* 2008); hidrocarburos como alcalenos de cadena larga (Mirzakhani y Nejad 2016); compuestos sulfonados o pigmentos azoicos muy utilizados para la coloración de artículos textiles y de cuero (Bell, Philp *et al.* 1998).

Todas estas características hacen que *Rhodococcus* sea una herramienta útil en procesos de biorremediación en combinación con otros organismos. Por ejemplo, *Rhodococcus erythropolis* CD106 aumenta considerablemente el proceso de fitorremediación de *Lolium perenne* en suelos contaminados por hidrocarburos derivados del petróleo (Płociniczak, Fic *et al.* 2017). Además, *Rhodococcus* se ha estudiado en procesos de biosorción de metales pesados (plomo, cadmio, cobre y zinc) en aguas y suelos contaminados con resultados prometedores (Vásquez, Botero *et al.* 2007, Emenike, Agamuthu *et al.* 2016).

El amplio rango de capacidades catabólicas de rhodococci se ha estudiado y en algunos caso incluso mejorado con fines comerciales. Los ejemplos más conocidos son la producción industrial de acrilamida por *Rhodococcus rhodochrous* J1 (Hughes, Armitage *et al.* 1998) y las transformaciones de nitrilo. Así por ejemplo, la cepa AJ270 de *R. erythropolis* convierte mezclas racémicas de fenilacetonitrilos y ciclopropanecarbonitrilos α-sustituidos en productos enantiopuros, como ácidos carboxílicos y amidas con rendimientos muy altos (Leng, Wang *et al.* 2009). Además de la producción de polímeros y amidas, los rhodococci también son capaces de sintetizar exopolisacáridos (Urai, Anzai *et al.* 2006, Urai, Yoshizaki *et al.* 2007), floculantes y agentes tensioactivos (Bell, Philp *et al.* 1998). Adicionalmente a todo lo mencionado, se ha visto que la cepa JCM 6824 de *R. erythropolis* produce auraquina RE, un antibiótico de quinolina, con un amplio y fuerte espectro antimicrobiano contra bacterias Gram-positivas (Kitagawa y Tamura 2008).

Rhodococci también se puede usar para la degradación de materiales de desecho, idealmente con el objetivo de producir compuestos útiles para la industria. Actualmente se están llevan a cabo varios estudios sobre el potencial que tiene el género *Rhodococcus* en la producción de lípidos para la obtención de biocombustibles a partir de aguas residuales o desechos de la

agricultura (Goswami, Namboodiri *et al.* 2017, He, Li *et al.* 2017, Le, Wells Jr *et al.* 2017). También, se ha demostrado que *Rhodococcus aetherivorans* IAR1 es capaz de usar tolueno (un contaminante volátil) como fuente de carbono con la producción simultánea de un plástico biodegradable, el poli (3-hidroxibutirato-co-3-hidroxivalerato) (Hori, Kobayashi *et al.* 2009).

2.1. Genomas y plásmidos de *Rhodococcus*

La versatilidad metabólica del género *Rhodococcus* se ve reflejada en su genoma (van der Geize y Dijkhuizen 2004), que codifica numerosas vías catabólicas para una gran variedad de compuestos químicos cuya degradación a su vez se canaliza en un menor número de rutas o vías "centrales" (Pieper y Seeger 2008). Las rutas metabólicas en rhodococci son complejas, cuidadosamente reguladas y al mismo tiempo flexibles, debido a la multiplicidad de genes y a los reordenamientos genómicos (Larkin, De Mot *et al.* 1998). La regulación transcripcional parece ser compleja en rhodococci tal como se evidencia por la presencia de genes que codifican proteínas reguladoras y por la dependencia de la transcripción de genes represores, inductores, y otros factores reguladores de la expresión génica (Komeda, Hori *et al.* 1996, Byrne, Russell *et al.* 2007, van der Geize, Yam *et al.* 2007, Takeo, Murakami *et al.* 2008, van der Geize, Hessels *et al.* 2008, Iida, Waki *et al.* 2009, Kendall, Burgess *et al.* 2010).

Gracias a la secuenciación masiva, se ha pasado de tener 4 genomas de *Rhodococcus* secuenciados en el 2012 a 191 genomas en la actualidad, y su número sigue creciendo. El tamaño de los genomas de *Rhodococcus* puede variar entre 3.9 Mb (*R. corynebacterioides* DSM 20151; GenBank: GCA_001646675,1) a 10 Mb de *R. koreensis JCM 10743* (GenBank: GCA_001894845.1). El DNA cromosómico y plasmídico de las cepas de *Rhodococcus* puede ser lineal o circular. Por ejemplo, tanto *R. jostii* RHA1 y *R. opacus* B4 tienen cromosomas lineales, mientras que *R. erythropolis* PR4 tiene un cromosoma circular. En la Tabla 1 se resumen las características principales de los genomas más representativos de *Rhodococcus* secuenciados hasta la fecha (último acceso, 29 de febrero del 2017), depositados en la base de datos del "National Center for Biotechnology Information Search database (NCBI)".

Varias cepas del género se caracterizan por presentar plásmidos y mega plásmidos. En *R. opacus* B4 se encuentran tres plásmidos circulares pKNR [110 kb; AP011118], pKNR01 [4.4 kb; AP011119] y pKNR02 [2,8 kb; AP011120] y dos plásmidos lineales pROB01 [560 kb; AP011116] y pROB02 [240 kb; AP011117].

Tabla 1. Lista de los genomas representativos de *Rhodococcus* secuenciados hasta la fecha.

Organismo	Ensamblaje	Tamaño (Mb)	GC%	Genes	Proteínas
R. corynebacterioides DSM 20151	GCA_001646675.1	3.9	70.2	3628	3408
R. kroppenstedtii DSM 44908	GCA_900111805.1	4.1	70.1	3799	3630
R. rhodnii LMG 5362	GCA_000389715.1	4.4	69.7	4260	3980
R. coprophilus NBRC 100603	GCA_001895045.1	4.6	66.9	4225	4075
R. gordoniae DSM 44689	GCA_001646655.1	4.8	68.0	4446	4244
R. marinonascens NBRC 14363	GCA_001894885.1	4.9	64.4	4663	4244
R. equi 103S	GCA_000196695.1	5.0	68.8	4614	4522
R. defluvii Ca11	GCA_000738775.1	5.1	68.7	1000	4488
R. rhodochrous DSM 43241	GCA_001646825.1	5.0	68.2	4758	4594
R. fascians D188	GCA_001620305.1	5.5	64.6	5138	5000
R. tukisamuensis NBRC 100609	GCA_001894985.1	5.5	69.9	4977	4768
R. kyotonensis KB10	GCA_001645385.1	5.5	65.2	3000	4893
R. pyridinivorans SB3094	GCA_000511305.1	5.6	67.8	5094	4916
R. kunmingensis DSM 45001	GCA_001646865.1	5.6	66.2	5282	4869
R. maanshanensis NBRC 100610	GCA_001894865.1	5.7	69.2	5153	4967
R. triatomae BKS 15-14	GCA_000341795.1	5.8	69.0	5275	5146
R. aetherivorans IcdP1	GCA_000982715.1	5.9	70.6	5326	5009
R. phenolicus DSM 44812	GCA_001646785.1	6.3	68.4	5851	5578
R. zopfii NBRC 100606	GCA_001895025.1	6.3	68.2	5887	5597
R. yunnanensis NBRC 103083	GCA_001895005.1	6.4	63.9	5860	5654
R. qingshengii BKS 20-40	GCA_000341815.1	6.6	62.4	6134	5978
R. globerulus NBRC 14531	GCA_001894805.1	6.7	61.7	6241	6053
R. erythropolis PR4	GCA_000010105.1	7.0	62.2	6428	6278
R. enclensis NIO-1009	GCA_001456965.1	7.5	62.3	7050	6833
R. imtechensis RKJ300	GCA_000260815.1	8.2	67.2	7661	7211
R. opacus B4	GCA_000010805.1	8.8	67.6	8048	7780
R. wratislaviensis IFP 2016	GCA_000325625.1	9.7	67.0	9389	8308
R. jostii RHA1	GCA_000014565.1	10.0	66.9	8816	8465
R. koreensis JCM 10743	GCA_001894845.1	10.0	67.4	9401	8782

Además, algunos de sus plásmidos son conocidos por su capacidad para conjugar. Por ejemplo, en *R. erythropolis* AN12 se ha encontrado dos plásmidos pREA400 [400 Kb; DQ489716.1] y pREA250 [250 Kb] con alta capacidad conjugativa (Yang, Lessard *et al.* 2007). Un resumen de algunos de los plásmidos identificados y cuya secuencia ha sido depositadas hasta la fecha, en el NCBI, se encuentran en la Tabla 2.

Tabla 2. Lista de los plásmidos identificados en el género *Rhodococcus* secuenciados hasta la fecha (último acceso, 29 de febrero del 2017).

Organismo	Plásmido	GenBank	Tamaño (Kb)	GC %	Proteínas	Genes	*PG
R. equi	p103	AF116907	80.60	64.6	64	64	-
	pREAT701	AP001204	80.61	64.6	64	64	-
	pVAPB1593	AM947676	79.25	64.7	68	72	4
	pVAPA1037	AM947677	80.61	64.6	65	73	8
	pVAPAMBE116	HM114217	83.10	64.6	69	69	-
R. erythropolis BG43	pRLCBG43	CP011296	240.12	60.5	207	225	18
	pRLLBG43	CP011297	261.53	61.3	258	273	15
	pRSLBG43	CP011298	29.46	58.7	22	24	2
R. erythropolis NI86/21	pFAJ2600	AF015088	5.93	60.8	8	8	-
R. erythropolis BD2	pBD2	AY223810	210.20	62.2	212	212	-
R. erythropolis DSM 8424	pRE8424	AB127588	5.98	54.7	6	6	-
R. erythropolis CCM2595	pRECF1	CP003762	90.22	62.5	98	101	3
R. erythropolis PR4	pREC1	AP008932	104.01	63.0	103	104	1
	pREC2	AP008933	3.637	62.2	5	5	-
	pREL1	AP008931	271.57	61.9	256	276	20
R. erythropolis R138	pCRE138	CP007256	91.72	60.7	81	89	8
	pLRE138	CP007257	477.91	60.4	411	439	28
R. rhodochrous B-276	pNC500	AB266604	7.63	65.7	8	8	-
R. jostii RHA1	pRHL1	CP000432	1123.08	65.1	949	1010	59
	pRHL2	CP000433	442.53	64.0	374	409	35
	pRHL3	CP000434	332.36	64.9	274	292	18
R. fascians D188	pFiD188	CP015236	197.39	61.8	164	172	8
	unnamed2	CP015237	164.72	64.3	155	163	8
R. opacus	pPDG1	CP008948	656.44	64.7	452	486	33
	pPDG2	CP008949	426.38	63.8	276	310	34
	pPDG3	CP008950	352.34	66.1	291	307	16
	pPDG4	CP008951	191.35	64.7	139	148	9
	pPDG5	CP008952	25.17	62.7	14	17	3
	pR1CP1	CP009112	885.38	64.1	687	758	68
	pR1CP2	CP009113	64.499	63.7	55	56	1
R. opacus B4	pKNR01	AP011119	4.36	64.5	7	7	-
	pKNR02	AP011120	2.77	63.8	-	-	-
	pROB01	AP011116	558.19	65.8	521	546	25
	pROB02	AP011117	244.99	6.4	196	211	15
	pKNR	AP011118	111.16	65.0	95	98	3
R. opacus PD630	1	CP003950	172.21	64.9	141	157	16
	2	CP003951	97.58	6.4	72	84	12
	3	CP003952	82.42	65.5	67	75	8
	4	CP003953	62.93	65.4	53	59	6
	5	CP003954	95.98	65.7	78	86	7
	6	CP003955	45.95	66.1	31	32	1
	7	CP003956	37.09	66.2	29	33	4
	8	CP003957	158.58	64.2	123	136	13
	9	CP003958	39.30	65.5	24	29	5

* PG: pseudogenes

Algunos rhodococci son conocidos por su inestabilidad genómica y reordenamientos. Así, encontramos ciertas cepas que poseen genes silenciosos que pueden expresarse después de reordenamientos genómicos y por plásmidos crípticos integrados en el cromosoma que pueden

cortarse, circularizarse y replicarse (Larkin, De Mot *et al.* 1998). Por ejemplo en *R. rhodochrous* NCIMB13064, pese a que originalmente es incapaz de crecer en compuestos aromáticos similares al naftaleno, se ha demostrado que algunas colonias adquirieron la capacidad de crecimiento gracias a reordenamientos genómicos asociados al plásmido pRTL1 (Kulakova, Reid *et al.* 1996). Los diferentes replicones presentes en una célula pueden tener diversos orígenes y pueden estabilizarse entre ellos, como se ha sugerido para *R. jostii* RHA1 (McLeod, Warren *et al.* 2006). Por otra parte, ciertas cepas de *R. fascians* pueden recombinarse (heterotalismo) o tienen la capacidad de integrar plásmidos en su genoma en secuencias aleatorias (Larkin, De Mot *et al.* 1998, Döhr, Paine *et al.* 2001). Todos estos ejemplos muestran la flexibilidad que tienen los genomas de *Rhodococcus*.

2.2. Herramientas para la manipulación genética de *Rhodococcus*

Gracias a la manipulación de flujos en una red metabólica, por medio de la modificación en la expresión de genes, se puede construir y mejorar cepas productoras de metabolitos de interés para la industria con el fin de obtener rendimientos más altos (Anesiadis, Cluett *et al.* 2008). En los últimos años el grupo de investigación de Ingeniería Metabólica del departamento de Bioquímica y Biología Molecular I, de la Universidad Complutense de Madrid, se ha centrado en la construcción de cepas de *Rhodococcus* para la conversión de fitoesteroles o colesterol en esteroides bioactivos.

Actualmente la gran disposición de secuencias génicas de *Rhodococcus* ha permitido el desarrollo de estrategias para la construcción de mutantes (van der Geize, Hessels *et al.* 2001, van der Geize y Dijkhuizen 2004, Von Bargen y Haas 2009) gracias a las cuales se ha logrado el análisis funcional de genes y proteínas. La deleción de genes sin marcador es particularmente útil dado que permite generación de mutantes con varias deleciones génicas (van der Geize, Hessels *et al.* 2001, van der Geize, Hessels *et al.* 2008). Como se mencionó anteriormente, la identificación de plásmidos de *Rhodococcus* ha fomentado el incremento de herramientas genéticas. De ese modo se han diseñado vectores lanzadera *Rhodococcus-E. coli*, que han sido usados con éxito para la generación de bibliotecas genómicas, estudios de expresión génica y de funcionalidad de genes (Singer y Finnerty 1988, Hashimoto, Nishiyama *et al.* 1992, Bigey, Grossiord *et al.* 1995, De Mot, Nagy *et al.* 1997, Duran 1998, Hirasawa, Ishii *et al.* 2001, Kostichka, Tao *et al.* 2003, Lessard, O'Brien *et al.* 2004, Nakashima y Tamura 2004, Matsui, Saeki *et al.* 2006). En varias especies de *Rhodococcus* se han empleado eficientemente sistemas de expresión usando los vectores pNit y pTIP (Nakashima y Tamura 2004) y vectores basado en

transposones (Sallam, Tamura *et al.* 2007), aunque no siempre son efectivos en todas las especies como es el caso de *R. ruber* Chol-4.

3. Los esteroides y esteroles

3.1. Definición y funciones biológicas

Los esteroides son compuestos orgánicos bioactivos de gran interés biológico. Se encuentran ampliamente distribuidos en la naturaleza y son indispensables para el funcionamiento de los organismos eucariotas. Las moléculas esteroideas abarcan hormonas, esteroles, ácidos biliares, vitamina D y diversos metabolitos secundarios, como las saponinas (Sigel, Sigel *et al.* 2007).

Los compuestos esteroideos están constituidos por 17 átomos de carbono formados por un ciclopentano-perhidrofenantreno (Figura 2A), estructura policíclica formada por 3 ciclohexanos (anillos A, B y C) y un ciclopentano (anillo D). Esta estructura puede presentar multitud de modificaciones, dando origen a una gran variedad de compuestos esteroideos. Las modificaciones más importantes son: i) presencia de un grupo hidroxilo o carbonilo en C3; ii) diferente grado de saturación de los anillos; iii) metilación en las posiciones C10 y C13 (da lugar a los carbonos C18 y C19) o iv) presencia de una cadena lateral alifática en C17 (JCBN 1989). La longitud de la cadena lateral y la presencia de diferentes grupos funcionales (en configuración α o β) determinan las características químicas, biológicas y fisiológicas de cada esteroide.

Se denomina esterol al núcleo esteroideo que posee un grupo hidroxilo en la posición C3 y una cadena lateral alifática de 8 átomos de carbono en la posición C17. Los esteroles dan fluidez y flexibilidad a las membranas eucariotas (Nakano, Motegi *et al.* 2007). Más del 25% de las membranas celulares pueden contener esteroles (Madigan, Martinko *et al.* 1997). Existen varios tipos de esteroles y su origen es diverso. En mamíferos el principal esterol es el colesterol (Figura 2B), en plantas los fitoesteroles (Figura 2D) (p. ej. β-sitosterol, campesterol y estigmasterol) y en hongos el ergosterol (Figura 2C).

Los esteroles son precursores de una gran variedad de compuestos con actividad biológica, tal es el caso de determinadas hormonas (p. ej. mineralocorticoides, glucocorticoides y hormonas sexuales) que funcionan como factores de transcripción para la regulación de la expresión génica. Estos factores actúan mediante la unión y la activación de moléculas receptoras cruciales para el control de los procesos biológicos en animales y seres humanos (Schmid y Urlacher 2007). Adicionalmente, los esteroles juegan un papel importante como parahormonas y

autohormonas, como agentes para la comunicación intercelular, como transductores de señales intracelulares y como reguladores reproductivos (Majewska 2007).

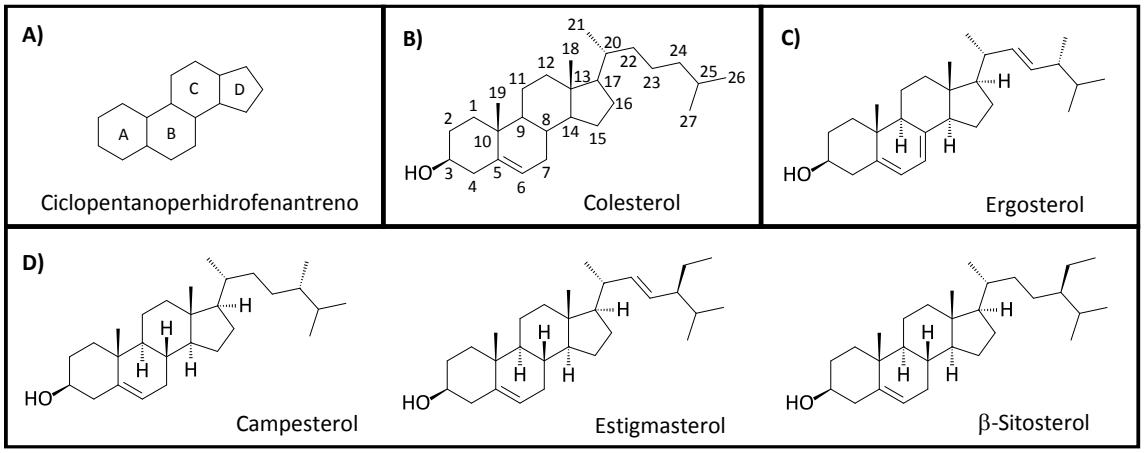


Figura 2. A) Representación esquemática del ciclopentano-perhidrofenantreno; **B)** representación del colesterol; **C)** representación del ergosterol; **D)** representación esquemática de tres fitosteroles (campesterol, β-sitosterol y estigmasterol).
Los sustituyentes en configuración α se representan con líneas discontinuas y los sustituyentes en configuración β con líneas sólidas, y la numeración corresponde a los átomos de carbono.

En mamíferos y hongos, los esteroides se forman a partir de la ruta del mevalonato. Esta ruta se inicia con la fusión del acetil-CoA a una acetoacetil-CoA, que eventualmente lleva a la generación de los precursores de los isoprenoides (isopentenil pirofosfato y dimetilalil pirofosfato). La ciclación de los isoprenoides da lugar al escualeno, que a su vez es convertido en lanosterol precursor básico para la formación de esteroides (Buhaescu y Izzedine 2007). En plantas superiores, el cicloartenol es un precursor de esteroides procedente de la ciclación del oxidoescualeno (Ohyama, Suzuki *et al.* 2009).

Los procariotas generalmente no son capaces de sintetizar esteroides, sin embargo sus membranas celulares pueden contener compuestos similares a los esteroides como pueden ser los terpenos pentacíclicos del tipo hopanoide que se sintetizan a partir del escualeno generado por una ruta independiente del mevalonato (Kannenberg y Poralla 1999, Bode, Zeggel *et al.* 2003). En esta ruta, las bacterias utilizan piruvato y gliceraldehído 3-fosfato para producir isopentenil pirofosfato y dimetilalil pirofosfato, que son los precursores del escualeno (Nakano, Motegi *et al.* 2007). No obstante, se han identificado ciertas especies bacterianas que pueden sintetizar esteroides y sus precursores. Algunos ejemplos son: los miembros del género *Mycoplasma* (*M. capricolum) (Dahl, Dahl et al. 1982)*, los actinomicetos (*Gordonia australis* cepa Acta 2299 y *Mycobacterium smegmatis*) (Lamb, Kelly *et al.* 1998, Schneider, Graf *et al.* 2008) y los planctomycetes (*Gemmata obscuriglobus)* (Pearson, Budin *et al.* 2003).

Por otro lado, los esteroides pueden ser degradados por bacterias capaces de utilizarlos como fuente de energía y de carbono para su crecimiento, tales como miembros del orden *Arthrobacter, Bacillus, Brevibacterium, Corynebacterium, Microbacterium, Mycobacterium, Nocardia, Protaminobacter, Serratia* y *Streptomyces*. Las bacterias del género *Rhodococcus*, orden Nocardia, son capaces de crecer en presencia de fitosteroles (p. ej. β-sitosterol, estigmasterol y campesterol), y esteroles animales (p. ej. colesterol, epicolesterol y 5α-colestanol) (Bell, Philp *et al.* 1998).

3.2. El colesterol: sus derivados e impacto ambiental

El colesterol (3-hidroxi-5,6-colesteno) es un alcohol policíclico de 27 átomos de carbono (Figura 2B), con un grupo hidroxilo en posición C3 (configuración β) (Hofmann y Hagey 2008). Es una molécula ampliamente distribuida en la naturaleza. Se encuentra formando parte de las membranas celulares (Lutton 1991) y abunda especialmente en el tejido nervioso y en el plasma como parte de las lipoproteínas (Bloch 1965, Tabas 2002).

En mamíferos, los esteroides -muchos derivados del colesterol- se dividen en varias clases basándose principalmente en su bioactividad: hormonas esteroideas C18 (estrógenos), C19 (andrógenos) y C21 (principalmente progestágenos), ácidos biliares y secosteroides. Estos últimos son derivados esteroideos, aunque el anillo B se encuentra abierto (Fahy, Subramaniam *et al.* 2005), como es el caso de la vitamina D3. Esta vitamina está relacionada con el transporte de calcio y tiene un rol fotoprotector (Lehmann y Meurer 2010). Los ácidos biliares derivados del colesterol -ya sea directamente (ácidos biliares primarios) o después de modificaciones que resultan en ácidos biliares secundarios- intervienen en la absorción de lípidos en el intestino, tienen efectos antimicrobiales y funcionan como moléculas señalizadoras (Lefebvre, Cariou *et al.* 2009).

Con la era industrial y los últimos avances tecnológicos se ha producido el vertido de numerosos compuestos contaminantes al medio ambiente (Jones, Voulvoulis *et al.* 2001). Estos vertidos contienen hidrocarburos, pesticidas, surfactantes y distintos metales, así como esteroles y esteroides de origen natural y sintético. Los vertidos proceden de actividades industriales como es el caso del lavado de la lana en el cual se usa lanolina; cuyo componente principal es el colestol (Poole 2004). Los efluentes de residuos industriales están expresamente reconocidos como altamente contaminantes por la legislación europea que regula su impacto ambiental (Directiva 85/377/CEE). En la actualidad, la mayor fuente de emisión de residuos contaminantes son los efluentes municipales, debido al uso que la población hace de tratamientos farmacológicos y

cosméticos (Chang, Wan *et al.* 2011). Los denominados "productos para el cuidado personal" (PPCPs) contienen compuestos esteroideos y colesterol, y sus derivados se consideran contaminantes de especial importancia debido a su impacto en los medios acuáticos (Aherne y Briggs 1989, Chang, Wan *et al.* 2011). Entre los productos de origen farmacéutico se encuentra el anticonceptivo sintético oral 17α- etinilestradiol, responsable -junto con los estrógenos 17α-estradiol y estrona- de causar la producción de vitelogenina (feminización) en peces macho (Sumpter y Jobling 1995).

El colesterol y sus derivados tienen gran estabilidad estructural, lo que les confiere una naturaleza recalcitrante que contribuye a su alto índice de persistencia. Es precisamente esta cualidad la que les dota de una naturaleza recalcitrante que contribuye a su alto índice de persistencia y permite que estos compuestos se utilicen como biomarcadores de referencia en algunos análisis de contaminación medioambiental (Guillette, Rooney *et al.* 1999, Veiga, Juste *et al.* 2005). La acumulación de estas moléculas en la naturaleza puede llegar a causar efectos muy perjudiciales sobre la salud humana y actualmente altera los ciclos biológicos de algunas especies animales que forman parte de estos ecosistemas.

4. Degradación bacteriana del colesterol y otros esteroides

Las primeras observaciones de microbios capaces de degradar colesterol se hicieron en 1913 (Söhngen 1913) y, desde entonces, una amplia variedad de bacterias ha sido identificadas. El estudio y conocimiento de las bases moleculares de las rutas catabólicas de estos compuestos podría permitir un uso racional de los procesos de síntesis de esteroides de interés industrial. Para ello, deben generarse enzimas con nuevas propiedades susceptibles de ser producidas, modificadas e incluso mejoradas mediante la ingeniería de proteínas. Sin embargo, muchas de las etapas enzimáticas responsables del catabolismo del colesterol y otros esteroides aún no son conocidas, y no se ha descrito ninguna ruta que explique la transformación completa desde colesterol a CO_2 y H_2O. Asimismo, la caracterización de las rutas de degradación de esteroles permitiría el empleo de procesos de ingeniería metabólica para la modificación, el control y la ampliación de las capacidades degradativas de diversas cepas bacterianas, con el objetivo final de su utilización como agentes biológicos descontaminantes.

Finalmente, es importante poner en evidencia la relación que existe entre los procesos de virulencia y la capacidad de degradar colesterol que existe en especies patógenas como *Mycobacterium tuberculosis* y *Rhodococcus equi* (Hondalus 1997, Brzostek, Dziadek *et al.* 2007).

La entrada del patógeno en la célula huésped es facilitada por la degradación total o parcial del colesterol presente en las membranas, que también puede provocar la lisis celular (Ahn y Sampson 2004). Además, el colesterol contribuye a la persistencia de *M. tuberculosis* ya que lo utiliza como fuente de carbono (Ouellet, Johnston *et al.* 2011).

4.1. Ruta de degradación del colesterol

Mediante la combinación de varios estudios realizados en diferentes cepas degradadoras de esteroles, se han podido determinar los pasos bioquímicos de algunas etapas del catabolismo del colesterol (van der Geize, Yam *et al.* 2007, Griffin, Gawronski *et al.* 2011, Thomas, Vanderven *et al.* 2011, Uhía, Galán *et al.* 2011, Uhía, Galan *et al.* 2011, Yam y Okamoto 2011, Uhía, Galán *et al.* 2012). Aunque las características fisiológicas de muchas de las enzimas implicadas aún no están definidas se ha podido identificar la existencia de dos eventos principales en el catabolismo de esteroles: la degradación de la cadena lateral en C17 y la ruptura del sistema de anillos (Figura 3 y Figura 4). Estos eventos pueden ocurrir simultáneamente o en un determinado orden (van der Geize, Hessels *et al.* 2002, Capyk, Casabon *et al.* 2011). Por ejemplo, en el caso de *Rhodococcus jostii* RHA1, se produce en primer lugar el ataque a la cadena lateral y posteriormente a la estructura de los anillos, si bien hay evidencias de cierta simultaneidad de ambos procesos en aquellos esteroides con una cadena lateral corta o parcialmente degradada (Rosłoniec *et al.*. 2009).

A continuación se exponen brevemente las diferentes etapas de esta ruta.

a) Transporte del colesterol

El transporte de esteroides al interior de la célula es uno de los pasos clave para su degradación. La agrupación de genes del catabolismo del colesterol incluye un sistema de transporte de esteroles codificado por el operón denominado Mce4, cuya función como transportador fue demostrada en *R. jostii* RHA1 y *Mycobacterium*. Las proteínas que constituyen el sistema Mce4 son: una ATPasa (Mce4G), dos subunidades de permeasas (SupA y SupB) y las proteínas Mce4A a Mce4I (Mohn, Van Der Geize *et al.* 2008). La mutación de los genes que codifican las proteínas supAB, mce4ABCDEF o mcc4IIJ impiden el crecimiento en colesterol, β-sitosterol, 5α-colestanol o 5α-colestanona. Por el contrario, el crecimiento en 4-androstene-3,17-diona, ácido cólico y progesterona no se ve afectado, lo que sugiere que existe una forma alternativa de entrada de esteroides. En la mayoría de proteínas del sistema Mce4 se han detectado péptidos señal en sus secuencias, lo que indica que son proteínas de secreción o de superficie (Mohn, Van Der Geize *et al.* 2008). En *M. smegmatis* ha sido descrito un sistema similar, requiriendo de 11 proteínas para ser funcional. Estas proteínas incluyen, al igual que en RHA1, una ATPasa (MceG), dos

subunidades de permeasas (YrbE4) y ocho proteínas de función aún desconocida (Mce4ABCDEFMas4AB). En *M. smegmatis* se ha descrito además un sistema pasivo de transporte de esteroides (García Fernández 2015), pero todavía queda mucho por conocer sobre sistemas de transporte de esteroides.

b) Transformación del colesterol en 4-colesten-3-ona

Una vez el colesterol ha entrado en la célula, se considera que el primer paso de degradación consiste en su oxidación por dos reacciones secuenciales: la oxidación de colesterol en su grupo hidroxilo C3 a 5-colesten-3-ona, seguida de una isomerización del doble enlace Δ^5 a Δ^3 formando 4-colesten-3-ona (colestenona). Este paso puede o no ser simultáneo a la degradación de la cadena lateral, y está mediado en algunos microorganismos por una colesterol oxidasa (ChOx) (Li, Vrielink *et al.* 1993, Yue, Kass *et al.* 1999, Coulombe, Yue *et al.* 2001, Navas, González-Zorn *et al.* 2001, Lario, Sampson *et al.* 2003, Brzostek, Dziadek *et al.* 2007, Drzyzga, Fernández de las Heras *et al.* 2011, Fernández de las Heras, Perera *et al.* 2014), mientras que en otros se trata de una 3β-hidroxiesteroide deshidrogenasa/isomerasa (3β-HSD) (Yang, Dubnau *et al.* 2007, Uhía, Galán *et al.* 2011, Yang, Gao *et al.* 2011). Estas enzimas tienen actividad 3β-hidroxiesteroide oxidasa o deshidrogenasa y una actividad Δ^5-3-cetoesteroide isomerasa. Se ha observado que, dependiendo de la cepa, las ChOx de rhodococci se inducen por esteroles vegetales, animales o por intermediarios esteroideos de la vía catabólica esteroidea (Kreit y Sampson 2009, Kumari y Shamsher 2015). En *R. erythropolis* CECT3014 se identificó una colesterol oxidasa de clase II (ChoG) que actúa tanto a nivel intracelular como extracelular. La mutación de este gen no impidió el crecimiento de la cepa en colesterol, pero sí perdió la capacidad de producir colesterol oxidasa extracelular, y la actividad de otras ChOx se vio afectada. Esto indica que otras enzimas estarían involucradas en el primer paso de degradación del colesterol en *R. erythropolis* (Fernández de las Heras, Perera *et al.* 2014). Curiosamente, en *R. jostii* cepa RHA1 se han identificado tres ChOx putativas y ninguna de ellas está relacionada con la degradación del colesterol debido a que en esta cepa no se ha detectado actividad ChOx en presencia de colesterol. Sin embargo, se ha demostrado que CYP125 (A14) realiza el paso inicial de degradación (van der Geize, Yam *et al.* 2007, Rosloniec, Wilbrink *et al.* 2009). Las células inducidas por colesterol de la cepa RHA1 revelaron una actividad 3βHSD en presencia de pregnenolona (un 3β-hidroxi esteroide con una cadena lateral C21 corta), pero no en presencia de colesterol (van der Geize, Yam *et al.* 2007). Por tanto, la función de las ChOx en el catabolismo de los esteroles parece ser dependiente de la cepa de *Rhodococcus*.

c) Oxidación de la cadena lateral del colesterol

Las primeras etapas de este proceso han sido estudiadas en *R. jostii* RHA1, *M. tuberculosis* y *M. smegmatis* (McLeod, Warren *et al.* 2006, van der Geize, Yam *et al.* 2007, Capyk, Kalscheuer *et al.* 2009, Rosloniec, Wilbrink *et al.* 2009, García-Fernández, Frank *et al.* 2013, Yang, Lu *et al.* 2015). Dependiendo de la cepa de *Rhodococcus*, después de la oxidación del colesterol a colestenona, la degradación de la cadena lateral puede ocurrir en diferentes puntos de la ruta. En este paso, el CYP125 (A14) cataliza la 26-hidroxilación del colesterol o de la 4-colesten-3-ona, dando lugar a la producción del ácido 26-hidroxi-4-colesten-3-ona (Figura 3) (Murohisa y Iida 1993, Murohisa y Iida 1993, Rosloniec, Wilbrink *et al.* 2009). En *R. jostii* RHA1 la degradación de la cadena lateral del colesterol y el catabolismo de los anillos esteroideos tienen lugar solo después de que se haya realizado la hidroxilación en C26 del esteroide por la actividad del CYP125 (A14). Por el contrario, la inactivación de CYP125A en una cepa mutante de *R. rhodochrous* DSM43269 no afectó la oxidación del anillo ni a la conversión del colesterol en 4-colesten-3-ona y 1,4-colestadien-3-ona, lo que demuestra que la oxidación del anillo no depende de la degradación de la cadena lateral (Rosloniec, Wilbrink *et al.* 2009). En *M. tuberculosis*, se mostró que el CYP125 hidroxila al colesterol y a la 4-colesten-3-ona, atacando inicialmente C26, C27 o ambos (Capyk, Kalscheuer *et al.* 2009). Una función similar la desempeña el CYP142 (Driscoll, McLean *et al.* 2010).

Después de la formación del ácido 3-oxo-colestenoico se cree que la cadena lateral es metabolizada mediante un proceso similar al de la β-oxidación de los ácidos grasos, y aunque las enzimas implicadas aún son desconocidas, actualmente se están identificando algunas de las proteínas que podrían intervenir en este proceso (Figura 3) (Wilbrink, Petrusma *et al.* 2011, Thomas y Sampson 2013, Yang, Lu *et al.* 2015). Por ejemplo, Nesbitt *et al.* (2010) publicó un estudio en el que la actividad tiolasa de β-cetoacetil-coenzima A (FadA5) en *M. tuberculosis* es necesaria para la producción de 4-androsteno-3,17-diona (AD) y 1,4-androstadien-3,17-diona (ADD) a partir del colesterol. En la cepa RHA1 se ha identificado un gen ortólogo que codifica FadA5, y se han propuesto otras enzimas putativas que pueden estar relacionadas con la β-oxidación tales como: las tiolasas Ltp3 y Ltp4, la ácido graso-CoA ligasa (FadD19), la acil-CoA deshidrogenasa (FadE26), la ácido graso-CoA sintetasa (FadD17), 17β- hidroxiesteroide deshidrogenasa (Hsd4A), la 2-enoil acil-CoA hidratasa (Hsd4B), la acil-CoA deshidrogenasa (FadE27) y la ácido graso CoA-hidratasa (EchA19) (van der Geize, Yam *et al.* 2007).

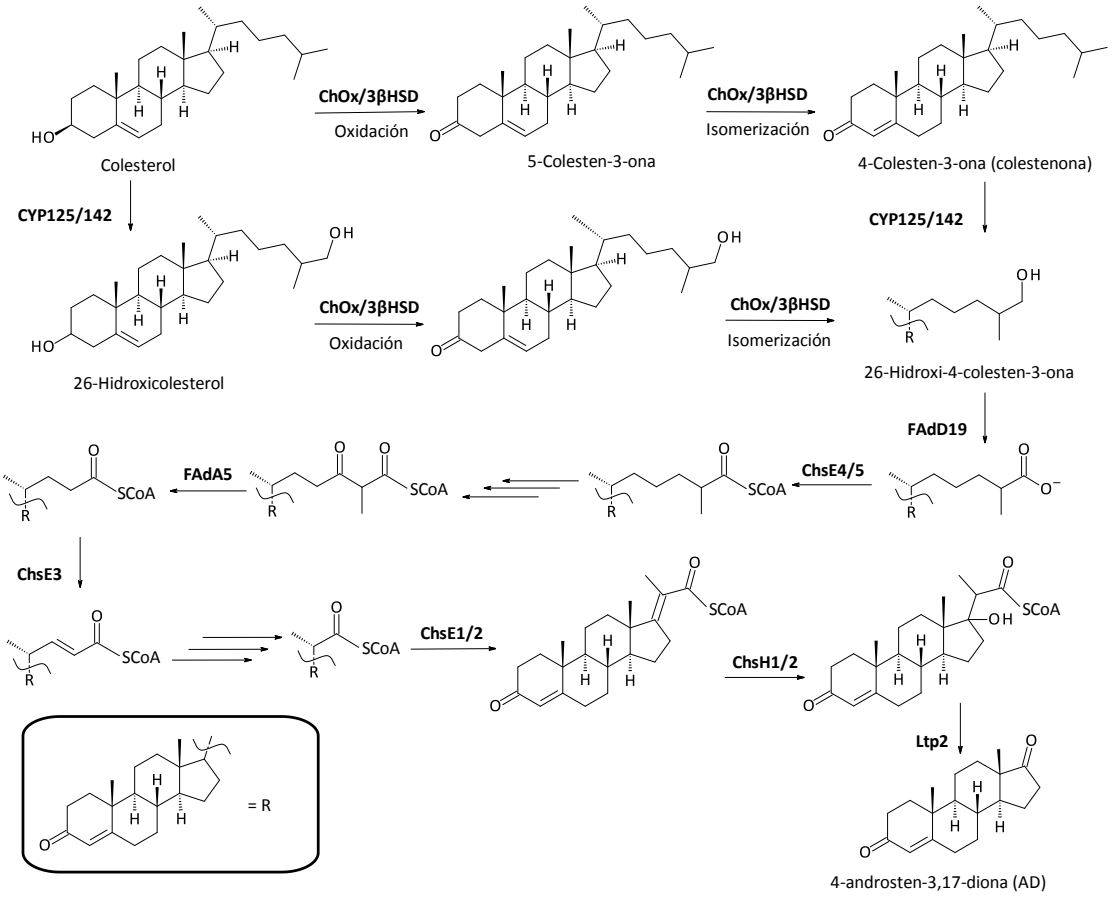


Figura 3. **Degradación de la cadena lateral del colesterol en Actinobacterias**. La presencia de múltiples flechas representa varios pasos enzimáticos. Antes o después de la conversión del colesterol en colestenona por la actividad de **ChOx** o **3βHSD**, el **CYP125** o el **CYP142** cataliza la oxidación en el C26, necesaria para el inicio de la degradación de la cadena lateral mediante un sistema similar a la β-oxidación.. Otras enzimas putativas que pueden estar relacionadas con la β-oxidación pero de las cuales no se sabe su función son: las tiolasas **Ltp3** y **Ltp4**, la acil-CoA deshidrogenasa (**FadE26**), la ácido graso-CoA sintetasa (**FadD17**), 17β-hidroxiesteroide deshidrogenasa (**Hsd4A**), la 2-enoil acil-CoA hidratasa (**Hsd4B**), la acil-CoA deshidrogenasa (**FadE27**) y la ácido graso CoA-hidratasa (**EchA19**). Este esquema de la degradación de la cadena lateral está basado en el propuesto por Bergstrand *et al.* (2016).

Wrońska *et al.* (2016) demostró que en *M. smegmatis* la enzima FadD19 es indispensable para el inicio de la degradación de esteroles C24 de cadena ramificada, pero no es indispensable para la degradación de los esteroles no ramificados (Wrońska, Brzostek *et al.* 2016). Además en *M. tuberculosis* se han identificado las enzimas ChsE1-ChsE2 que catalizan la deshidrogenación de 3-oxo-4-pregneno-20-carboxil-CoA (3-OPC-CoA) (Yang, Guja *et al.* 2014), las ChsE4-ChsE5 que catalizan la oxidación de 3-oxo-colest-4-en-26-oil CoA, 3-oxo-col-4-en-24-oil CoA así como 3-oxo-4-pregnen-20-carboxil-CoA, en el primer ciclo de β-oxidación de la cadena lateral de colesterol, lo que demuestra una redundancia funcional (Figura 3). Otra enzima, la ChsE3, cataliza específicamente la oxidación de 3-oxo -col-4-en-24-oil-CoA en el segundo ciclo de β-

oxidación (Figura 3) (Yang, Lu *et al.* 2015). Se han identificado algunos productos de estos procesos de β-oxidación, como por ejemplo dos tipos de intermediarios en *Mycobacterium*: los esteroides C19 (p. ej. AD, ADD) y los esteroides C22 (23,24-bisnocolenoicos, HBCs) (p. ej. 22-hidroxi-23,24-bisnorcol-4-ene-3-ona, 4-HBC; 22-hidroxi-23,24-bisnorcol-1,4-dien-3-ona, 1,4-HBC; 9,22-dihidroxi-23,24-bisnorcol-4-ene-3-ona, 9OHHBC). Esto indica la posibilidad de dos rutas de degradación de esteroides: la ruta por AD y la ruta por HBC (Xu, Liu *et al.* 2016). Recientemente, se ha demostrado que la enzima Hsd4A de *Mycobacterium neoaurum* ATCC 25795 tiene actividad 17β-hidroxiesteroide deshidrogenasa y β-hidroxiacil-CoA deshidrogrenasa, indispensable en la ruta de los esteroides 23,24-bisnocolenoicos (Xu, Liu *et al.* 2016).

d) Ruta central degradación del colesterol

Tras la oxidación del colesterol por ChOx o la 3βHSD, se inicia la ruptura de los anillos esteroideos que puede ser o no concomitante con la degradación de la cadena lateral (Figura 4). En *Mycobacterium*, esta ruta parece ser común tanto para esteroides C19 como para C22 (Xu, Liu *et al.* 2016) aunque recientemente se ha publicado una ruta alternativa para los esteroides C19 vía el regulador padR (Fernández-Cabezón, García-Fernández *et al.* 2017). En *R. erythropolis* se ha visto que la ruptura de la estructura policíclica empieza con la apertura del anillo B por la hidroxilación de C9 catalizada por una 3-cetoesteroide 9α-hidroxilasa (KshAB) (van der Geize, Hessels *et al.* 2002, van der Geize, Yam *et al.* 2007, Petrusma, Dijkhuizen *et al.* 2009, Petrusma, Hessels *et al.* 2011, Petrusma, Dijkhuizen *et al.* 2012) y la deshidrogenación del anillo A en C1-C2 por una 3-cetoesteroide-Δ^1 –deshidrogenasa (KstD) (van der Geize, Hessels *et al.* 2000, van der Geize, Hessels *et al.* 2001, van der Geize, Hessels *et al.* 2002, van der Geize y Dijkhuizen 2004, Knol 2011, Fernández de las Heras, van der Geize *et al.* 2012). Las KstDs son enzimas que convierten el AD en ADD o el 9α-hidroxi-4-androsten-3,17-diona (9OH-AD) en 9α-hidroxi-1,4-androstadien-3,17-diona (9OH-ADD), formando un doble enlace entre C1 y C2. Las KshABs añaden un grupo hidroxilo en el C9, dando lugar a la formación de 9OH-AD y 9OH-ADD a partir de AD o ADD respectivamente. La actividad combinada de ambas enzimas da como resultado el 9OH-ADD. Este último compuesto es químicamente inestable y sufre una hidrólisis espontánea que da lugar a la 3-hidroxi-9,10-secoandrosta-1,3,5(10)-trien-9,17-diona (3-HSA). En *R. jostii* RHA1. *M. tuberculosis* y *C. testosteroni*, la hidroxilación subsecuente del 3-HSA está mediada por oxigenasas de dos componentes (HsaAB) (Horinouchi, Hayashi *et al.* 2005, van der Geize, Yam *et al.* 2007, Dresen, Lin *et al.* 2010). Esta actividad da como resultado un derivado catecólico, la 3,4-dihidroxi-9,10-secoandrosta-1,3,5(10)-trien-9,17-diona (3,4-DHSA). En *R. jostii* RHA1 (van

der Geize, Yam *et al.* 2007) y en *M. tuberculosis* (Yam, D'Angelo *et al.* 2009), el anillo A de la 3,4-DHSA se abre por una extradiol dioxigenasa de tipo meta, la HsaC, originando el ácido 4,5,9,10-diseco-3-hidroxi-5,9,17-trioxoandrosta-1(10),2-dien-4-oico (4,9-DSHA). Este compuesto es hidrolizado por la HsaD en *R. jostii* RHA1 y *M. tuberculosis* (Larkin, Kulakov *et al.* 2005, van der Geize, Yam *et al.* 2007), originando los ácidos 2-hidroxi-2,4-hexadienoico (2-HHD) y 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oico (HIP) (Chipley, Dreyfuss *et al.* 1975). El proceso está ilustrado en la Figura 4.

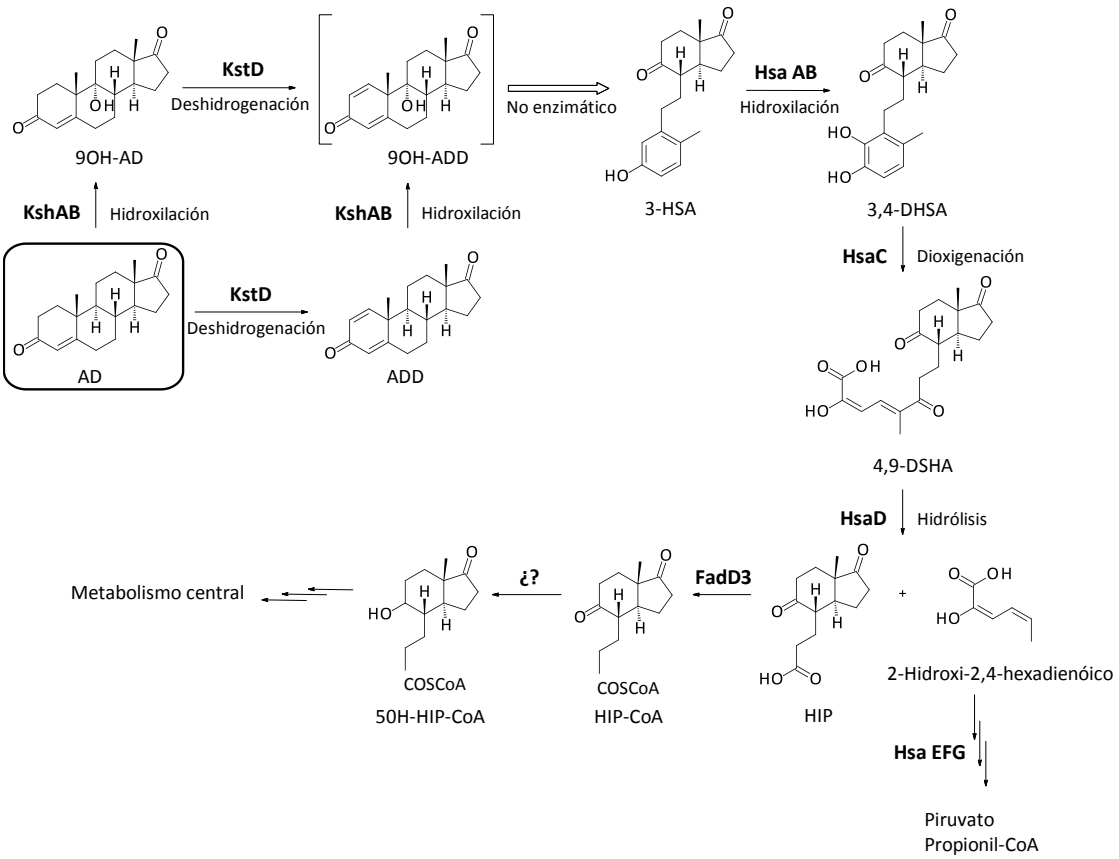


Figura 4. **Ruta propuesta de degradación del colesterol en bacterias aerobias**. Esquema de la rotura del sistema de anillos esteroideos una vez degradada la cadena lateral (ver Figura 3). La presencia de flechas múltiples indica un proceso multienzimático. Abreviaturas de los compuestos: **AD**: 4-androsten-3,17-diona; **ADD**: 1,4-androstadien-3,17-diona; **9OH-AD:** 9-hidroxi-4-androsten-3,17-diona; **9OH-ADD**: 9-hidroxi-1,4-androstadien-3,17-diona; **3-HSA**: 3-hidroxi-9,10-secoandrosta-1,3,5(10)-trien-9,17-diona; **3,4-HSA**: 3,4-dihidroxi-9,10-secoandrosta-1,3,5(10)-trien-9,17-diona; **4,9-DSHA**: ácido 4,5-9,10-diseco-3-hidroxi-5,9,17-trioxoandrosta-1(10),2-dien-4-oico; **HIP**: ácido 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oico; **HIP-CoA**: derivado CoA del HIP; **5OH-HIP-CoA**: derivado hidroxilado del HIP-CoA. Abreviaturas de las enzimas: **KstD**: 3-cetoesteroide-Δ^1-deshidrogenasa; **KshAB**: 3-cetoesteroide-9α-hidroxilasa; **HsaAB**: 3-hidroxi-9,10-secoandrosta-1,3,5(10)-trien-9,17-diona 4-hidroxilasa; **HsaC**: 3,4-dihidroxi-9,10-secoandrosta-1,3,5(10)-trien-9,17-diona dioxigenasa; **HsaD**: ácido 4,5-9,10-diseco-3-hidroxi-5,9,17-trioxoandrosta-1(10),2-dien-4-oico hidroxilasa; **HsaE**: 2-hidroxipentadienoato hidratasa; **HsaF**: 4-hidroxi-2-cetovalerato aldolasa; **HsaG**: acetaldehído deshidrogenasa; **FadD3**: HIP-CoA transferasa.

e) Ruta baja de degradación del colesterol

Se conoce muy poco sobre esta ruta, pero probablemente el ácido 2-hidroxi-2,4-hexadienoico (2-HHD) es metabolizado por las enzimas hidratasa (HsaE), aldolasa (HsaF) y acetaldehído deshidrogenasa (HsaG) dando lugar al ácido pirúvico y el propionil-CoA, metabolitos que entrarían en las rutas centrales de degradación (Kieslich 1985, Horinouchi, Hayashi *et al.* 2003, van der Geize, Yam *et al.* 2007). A pesar de que estas enzimas no han sido caracterizadas, presentan una elevada similitud con las proteínas TesE, TesG y TesF de *Comamonas testosteroni* implicadas en catalizar reacciones en la ruta de degradación de la testosterona (Horinouchi, Hayashi *et al.* 2005).

Se cree que el metabolismo del HIP podría iniciarse mediante la degradación de su cadena lateral a través de un proceso parecido a la β-oxidación (Lee y Sih 1967) cuyo primer paso bioquímico consistiría en la activación de HIP con CoA. La enzima FadD3 (acil-CoA sintetasa) ha sido descrita como la responsable de dicha reacción en *R. jostii* RHA1 (Casabon, Crowe *et al.* 2013).

A continuación actuaría una HIP-CoA reductasa que transformaría el grupo ceto del carbono C5 en un hidroxilo y el derivado hidroxilado sería sustrato de FadE30 (acil-CoA deshidrogenasa) que iniciaría el ciclo de β-oxidación (Van der Geize, Grommen *et al.* 2011). El resto de las etapas de degradación de los anillos C y D, que llevarían a la producción de metabolitos centrales, todavía son objetos de estudio.

5. Regulación del metabolismo del colesterol

Poco se conoce sobre la regulación de los genes del metabolismo de esteroides. Hasta la fecha, tanto en *M. smegmatis* como en *R. jostii* RHA1, se han identificado dos reguladores transcripcionales: KstR y KstR2, ambos pertenecientes a la familia TetR (Kendall, Withers *et al.* 2007, Kendall, Burgess *et al.* 2010, Uhía, Galán *et al.* 2011). Estos reguladores actúan como represores de los genes de la ruta: KstR controla la expresión de 83 genes en *M. smegmatis* y 74 genes en *M. tuberculosis* relacionados con el metabolismo del colesterol, concretamente en la parte alta y central de la ruta (Kendall, Withers *et al.* 2007, Nesbitt, Yang *et al.* 2010). El represor KstR2 controla la expresión de 15 genes relacionados con la parte baja de la ruta tanto en *M. smegmatis* como en *M. tuberculosis* (Kendall, Burgess *et al.* 2010, García-Fernández, Galán *et al.* 2015). Mediante estudios de las regiones promotoras de los genes regulados por KstR y KstR2, se han detectado secuencias palindrómicas que actúan como secuencias reguladoras. La secuencia operadora para KstR es TnnAACnnGTTnnA y para KstR2 es AAnCAAGnnCTTGnTT

(Kendall, Burgess *et al.* 2010, García-Fernández, Galán *et al.* 2015). En *M. smegmatis*, se ha demostrado que los ácidos 4-colesten-3-ona-26-oico o el 3-oxo-colestenoico actúan como efectores del represor KstR permitiendo la expresión de los genes implicados en el metabolismo del colesterol (García-Fernández, Medrano *et al.* 2014). Por otra parte, se ha descrito en *M. tuberculosis* y en *R. jostii* RhA1 que los genes regulados por KstR2 se inducen en presencia de colesterol, HIP y otros derivados de la ruta. Además se ha demostrado que HIP-CoA actúa como un efector de KstR2 (Casabon, Zhu *et al.* 2013).

6. Enzimas objeto de estudio en este trabajo: 3-cetoesteroide-Δ^1-deshidrogenasas (KstD) y 3-cetoesteroide-9α-hidroxilasa (KshAB)

6.1. Las 3-cetoesteroide-Δ^1-deshidrogenasas (KstD)

Las 3-cetoesteroide-Δ^1-deshidrogenasas [Δ^1-KSTD; 3-oxosteroid 1-dehydrogenase; EC 1.3.99.4] en adelante KstD, son flavoenzimas que catalizan la Δ^1-deshidrogenación de los 3-cetoesteroides mediante la desprotonación de C1 (α) y C2 (β) del anillo A, con la consecuente formación de un doble enlace (Itagaki, Hatta *et al.* 1990). En bacterias, las KstDs junto con las KshABs son claves en la apertura del anillo B de los esteroides. Se ha demostrado que las KstDs catalizan tanto la conversión de AD en ADD, como la conversión de 9OH-AD al compuesto químicamente inestable 9OH-ADD (Figura 4). Algunos productos de la 3-cetoesteroide deshidrogenasa son los intermediarios más comercializados para la producción de esteroides farmacéuticos tal como la 5-androstan-3.17-diona (1-(5α-AD)) y el ADD (Dijkstra. 2016 #424). Las 3-cetoesteroide-Δ^1-deshidrogenasas se encuentran en muchas bacterias degradadoras de esteroides, y de hecho se han identificado y caracterizado genes homólogos de *kstD* en: *C. testosteroni* (Plesiat, Grandguillot *et al.* 1991), *R. erythropolis* SQ1 (van der Geize, Hessels *et al.* 2000, van der Geize, Hessels *et al.* 2001, Knol, Bodewits *et al.* 2008), *Rhodococcus ruber* Chol-4 (Fernández de las Heras, van der Geize *et al.* 2012), *M. smegmatis* (Brzostek, Sliwinski *et al.* 2005), *Nocardia corallina* (Itagaki, Wakabayashi *et al.* 1990), *R. rhodochrous (Liu, Shen et al. 2016,* Morii, Fujii *et al.* 1998), entre otros. En el genoma de *R. jostii* RHA1 se han identificado 6 KstDs putativas (Mathieu, Schloendorn *et al.* 2008), en *R. ruber* Chol-4 y en *R. erythropolis* SQ1 se han encontrado al menos 3 genes *kstD* (van der Geize, Hessels *et al.* 2001, Knol 2011, Fernández de las Heras, van der Geize *et al.* 2012). La presencia de tantas isoenzimas KstDs en rhodococci todavía está en estudio, pero se ha demostrado que la afinidad por el sustrato varía según la isoenzima (Knol, Bodewits *et al.* 2008, Liu, Shen *et al.* 2016). Además, cabe mencionar que las KstDs pueden estar relacionadas con otras rutas metabólicas como por ejemplo la KstD de *A.*

fumigatus CICC 40167 ($KstD_F$) que está involucrada en la biosíntesis del antibiótico fusadina (Chen, Wang *et al.* 2012). La presencia de varias isoformas de una misma enzima, hace que en algunos casos sea necesaria la mutagénesis secuencial de más de un gen con el fin de interrumpir su actividad.

6.1.1. Localización celular de las KstDs

Se conoce poco sobre la localización celular de las KstDs. En *R. erythropolis* IMET 7030 se ha localizado una KstD por marcaje inmunoquímico. La fracción mayoritaria se detectó en la región periférica de la membrana citoplasmática, pero también estaba presente en canales que conectan el citoplasma con la superficie celular (Wagner, Atrat *et al.* 1992). Las 3 enzimas KstD de *R. erythropolis* SQ1 se han expresado heterólogamente en *E. coli*, y todas ellas se encuentran como proteínas solubles en los extractos celulares (Wagner, Atrat *et al.* 1992, van der Geize, Hessels *et al.* 2000, van der Geize, Hessels *et al.* 2002, Knol, Bodewits *et al.* 2008). De la misma forma, esta actividad también es mayoritaria en las fracciones citosólicas en el caso de *Mycobacterium sp.* VKM Ac-1817D (Sukhodolskaya, Nikolayeva *et al.* 2007). No obstante, se ha observado que al expresar el gen *kstD* de *C. testosteroni* en *E. coli*, la fracción con proteína activa se asociaba a la membrana interna (Plesiat, Grandguillot *et al.* 1991).

6.1.2. Regulación de las KstDs

En cuanto a la regulación de la expresión génica de las KstDs existen estudios de inducción en presencia de diferentes fuentes de carbono. Por ejemplo, *R. erythropolis* IMET 7030 aumenta sus niveles de KstD al inducir con 17α-metil-testosterona (Wagner, Atrat *et al.* 1992). Mediante estudios de microarrays en *R. jostii* RHA1 se ha visto que *kstD* (*ro04532*) se induce por colesterol, mientras que otras *kstDs* también presentes en su genoma se inducen en presencia de 7-cetocolesterol (Mathieu, Mohn *et al.* 2010). En *Gordonia neofelifaecis* NRRL B-59395 se han detectado 5 KstDs. Los análisis transcripcionales revelaron que la expresión de *kstD1, kstD3* y *kstD4* aumentó en el cultivo en presencia de colesterol o AD. Por otro lado, *kstD2* y *kstD5* se indujeron solo en presencia de colesterol (Zhang, Ren *et al.* 2015). En *Rhodococcus rhodochrous* DSM43269 se han identificado 3 KstDs, de las cuales, sólo los genes *kstD1* y *kstD2* se inducen en presencia de AD (Liu, Shen *et al.* 2016). Estos resultados apuntan a un mecanismo finamente regulado que depende de la fuente de energía y carbono disponible.

6.1.3. Estructura y mecanismo de acción de las KstDs

La estructura de las KstDs y su mecanismo de acción se han podido determinar gracias a los análisis cristalográficos de KstD1 de *R. erythropolis* SQ1 y KstD4 de *R. jostii* RHA1 (Rohman, van Oosterwijk *et al.* 2012, Rohman, van Oosterwijk *et al.* 2013, Dijkstra, van Oosterwijk *et al.* 2016). KstD1 cristaliza como una unidad asimétrica conformada por ocho monómeros. Cada monómero tiene una forma elongada y presenta dos dominios: uno de unión a FAD y otro catalítico (Figura 5A). Los monómeros forman dos tetrámeros independientes y similares entre sí (Figura 5B) (Rohman, van Oosterwijk *et al.* 2013). El dominio FAD consta de dos segmentos polipetídicos separados (residuos 3-278 y 449-510) que forman dos subdominios FAD-A y FAD-B. FAD-A tiene un plegamiento de Rossman típico de las enzimas de unión a FAD. El dominio catalítico (residuos 279-448) se encuentra entre la α9 y β19 del subdominio FAD-A (Rohman, van Oosterwijk *et al.* 2013). El dominio FAD tiene una secuencia consenso de unión que se encuentra altamente conservada: XhXhGXGXXGXXXhXXhX$_7$hXhE(D), donde X es cualquier residuo y h es un residuo hidrofóbico (Dym y Eisenberg 2001).

La estructura del sitio activo tiene una cavidad similar a la forma de un bolsillo y está formada por los residuos hidrofóbicos del domino catalítico y el domino de unión a FAD. La naturaleza apolar de los residuos del bolsillo está conservada entre KstDs de diferentes bacterias, al igual que los residuos de interacción con el sustrato (Tyr^{119}, Tyr^{319}, Tyr^{487} y Gly^{491}) (Rohman, van Oosterwijk *et al.* 2013, Dijkstra, van Oosterwijk *et al.* 2016). Para determinar la importancia de estos residuos, las tirosinas fueron mutadas reemplazándolas por fenilalaninas. Estos mutantes mostraron una reducción importante de la actividad enzimática. La mutación de los residuos dio como resultado la completa supresión de la actividad deshidrogenasa para el caso de la Tyr^{318} y una reducción del 99% en el caso de la Tyr^{119}. La mutación de la Tyr^{487} dio como resultado una actividad de un 2.6% con respecto al fenotipo silvestre (Rohman, van Oosterwijk *et al.* 2013, Dijkstra, van Oosterwijk *et al.* 2016).

Por otra parte, en el caso de la KstD2 de *R. erythropolis,* se ha determinado, mediante mutagénesis por irradiación con luz UV, que los residuos Thr^{503} y Ser^{325} son esenciales para la actividad de esta proteína (van der Geize, Hessels *et al.* 2001). La Thr^{503} forma parte de una región altamente conservada en todas las KstDs de *Rhodococcus*. En cambio, la Ser^{325} solo está parcialmente conservada entre las proteínas KstD2.

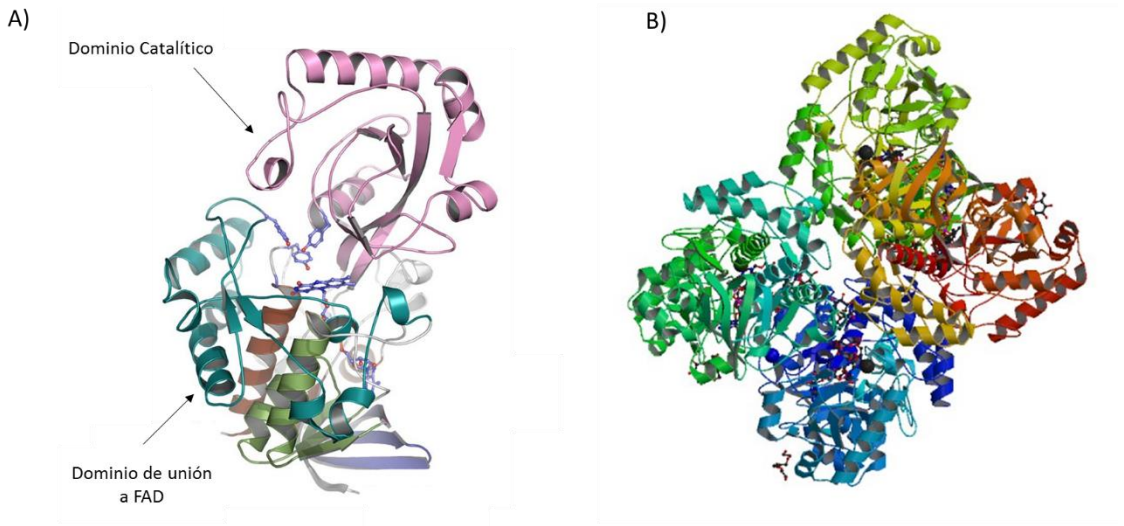


Figura 5. Estructura cristalográfica de KstD1 de *R. erythropolis* SQ1 (Protein Data Bank (PDB): 4C3X). **A)** Monómero publicado por Rhoman *et al.* (2013). El dominio de unión a FAD está representado en tonos de verde y marrón, el dominio catalítico en rosa. Los residuos de unión a FAD y al sustrato del dominio catalítico están resaltados en azul. **B)** Octámero visto desde un eje simétrico (figura del ensamblaje biológico tomado del PDB).

El mecanismo catalítico de deshidrogenación de las KstDs (Figura 6) se inicia con la unión del grupo hidroxilo de la Tyr^{487} y la amina de la cadena principal de la Gly^{491} al grupo ceto C3 del sustrato, promoviendo la tautomerización del ceto-enol y la desestabilización de los átomos de hidrógeno C2. La Tyr^{318} -cuyo carácter básico se ve reforzado por un enlace de hidrógeno con la Tyr^{119}- sustrae el hidrógeno del átomo de C2 en forma de protón (Rohman, van Oosterwijk *et al.* 2013, Dijkstra, van Oosterwijk *et al.* 2016). El enolato resultante puede experimentar intercambio de hidrógenos en el átomo de C2 con el disolvente a través de la Tyr^{318}. Tras la migración de la carga negativa del sustrato al átomo C1, ocurre la transferencia como un ion hidruro del hidrógeno del átomo C1 al átomo N5 del anillo isoaloxazina de FAD. Esto da lugar a la formación de un doble enlace entre los átomos C1 y C2. La carga negativa en el FADH se estabiliza por el momento dipolar cerca de la α-hélice (Rohman, van Oosterwijk *et al.* 2013, Dijkstra, van Oosterwijk *et al.* 2016).

Con el fin de determinar la actividad de estas enzimas se han utilizado diferentes aceptores de electrones artificiales en ensayos cinéticos, como el DCPIP (2,6-diclorofenolindofenol), el PMS (metosulfato de fenazina) más NBT (nitroazul de tetrazolio), el azul de Wurster (TMPD: N,N,N',N'-tetrametil-*p*-fenilendiamina) y la vitamina K3 (Abul-Hajj 1978, Plesiat, Grandguillot *et al.* 1991, van der Geize, Hessels *et al.* 2000).

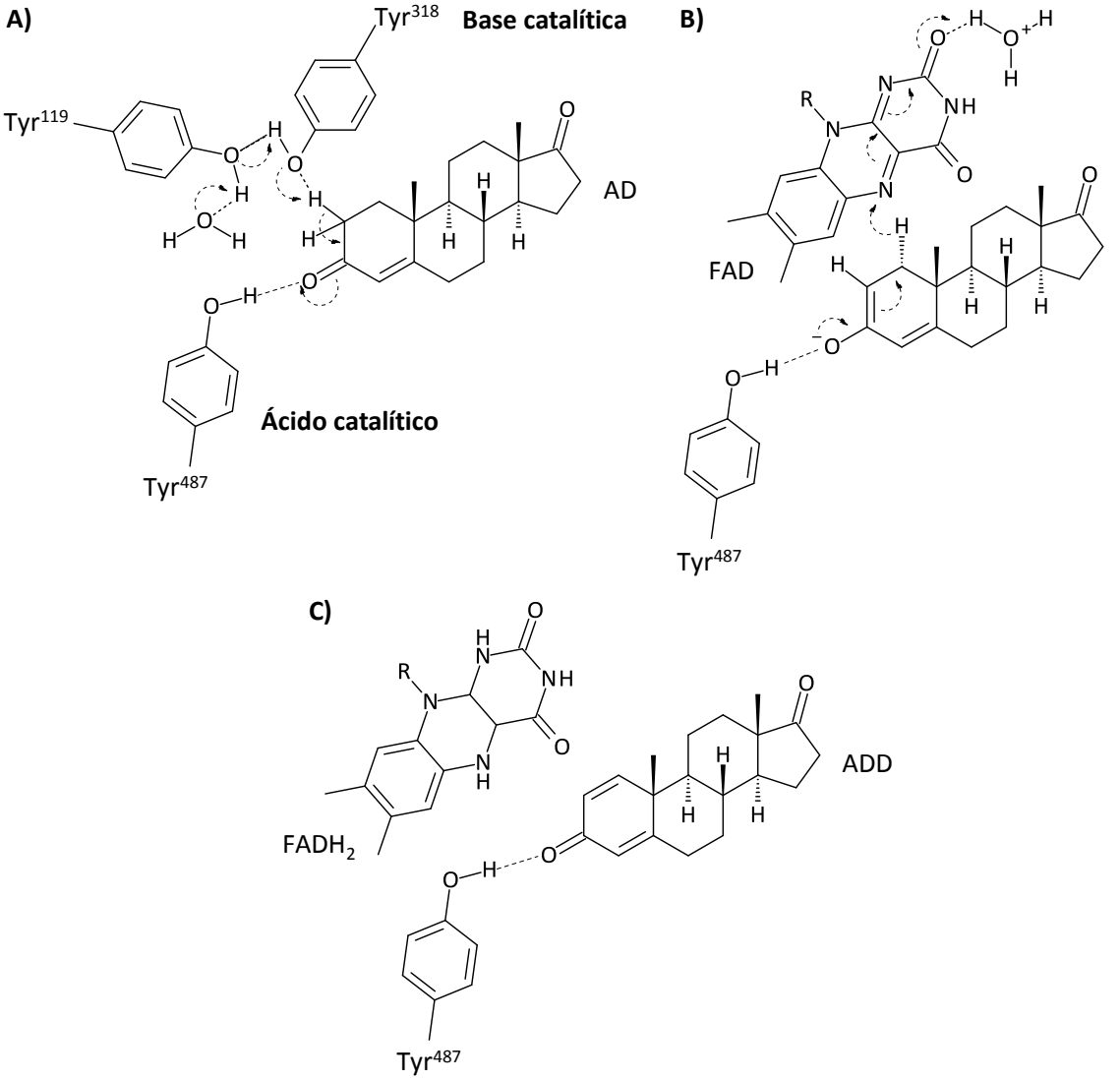


Figura 6. Mecanismo de deshidrogenación propuesto para KstD1 de *R. erythropolis* SQ1 (Dijkstra, van Oosterwijk *et al.* 2016). **A)** La interacción del grupo ceto del sustrato AD con el hidroxilo de la **Tyr^{487}** (ácido catalítico) desencadena la tautomerización e inestabilidad de los átomos de hidrógeno del C1 y C2 en el AD. La **Tyr^{318}** (base catalítica) toma un protón del C2, y la **Tyr^{119}** permite la transferencia del protón al solvente. **B)** El FAD sustrae un ion hidruro del C1 del sustrato y -posteriormente a una reordenación estructural- se da la formación de un doble enlace entre C1-C2 en el sustrato, con la consecuente formación de FADH el cual a su vez interactuará con el protón liberado inicialmente al solvente. **C)** El producto formado (ADD) se libera del centro activo y se forma una molécula de $FADH_2$ tras un tautomerismo ceto-enol.

6.1.4. Tipos de KstDs y sustratos

Las KstDs pueden deshidrogenar una gran variedad de 3-cetoesteroides, pero no 3-hidroxiesteroides, y generalmente tienen preferencia por los sustratos no saturados en C4-C5 (p. ej. AD, cortisol, progesterona) (Zhang, Ren *et al.* 2015). Se ha demostrado que las KstDs presentan un amplio rango metabólico. Varios ejemplos de sustratos y productos se pueden encontrar en la Tabla 3 (Itagaki, Wakabayashi *et al.* 1990, Knol, Bodewits *et al.* 2008). Gracias a

la elaboración de árboles filogenéticos, se han podido distinguir al menos cuatro grupos de KstDs (Figura 7) (Knol, Bodewits *et al.* 2008, Fernández de las Heras, van der Geize *et al.* 2012).

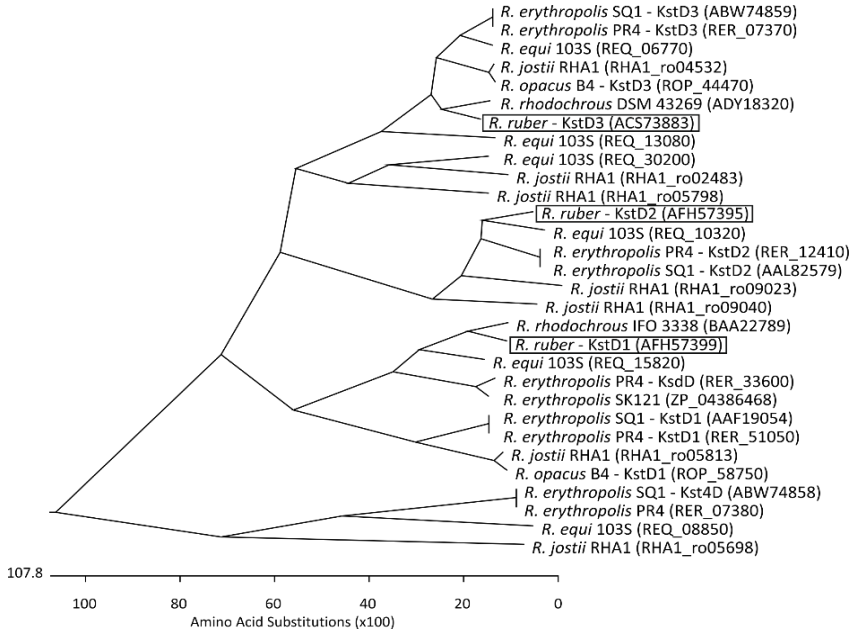


Figura 7. Árbol filogenético de KstDs de *Rhodococcus*, publicado por Fernández-de las Heras *et al.* (2012). Las enzimas de *R. ruber* Chol-4 están resaltadas en los recuadros.

Se han caracterizado 3 KstDs en *R. erythropolis* SQ1 y en *R. ruber* Chol-4, cada una perteneciente a grupos diferentes. Las enzimas KstD1 y KstD2 de *R. erythropolis* SQ1 tienen preferencia por sustratos que presentan estructuras 4-en-3-ceto, como es el caso de AD y 9OH-AD. Sin embargo, KstD3 tiene una clara preferencia por los 3-cetoesteroides con el anillo A saturado, mostrando los mayores niveles de actividad con 5α-AD (5α-androstan-3,17-diona) y 5α-T (5α-testosterona) (Knol, Bodewits *et al.* 2008).

En bacterias degradadoras de colesterol como *C. testosteroni*, *R. jostii* o *R. rhodochrous*, también se ha descrito la presencia de otro tipo de 3-cetoesteroide deshidrogenasas, conocidas como 3-cetosteroide-Δ^4-deshidrogenasas [Δ4-(5α)-KSTD; EC 1.3.99.5].

Estas enzimas catalizan la desprotonación de C4(β) y C5(α) de los 3-cetoesteroides (Florin, Kohler *et al.* 1996, Knol, Bodewits *et al.* 2008, Penfield, Worrall *et al.* 2014).En *R. jostii* RH1, se ha caracterizado una Δ^4-(5α)-KSTD similar a las KstD puesto que también participa en la degradación del anillo esteroideo, además de presentar un mecanismo de acción muy parecido (Dijkstra, van Oosterwijk *et al.* 2016, van Oosterwijk, Knol *et al.* 2011, van Oosterwijk, Knol *et al.* 2012).

Tabla 3. Reacciones enzimáticas catalizadas por KstDs en diferentes organismos.

Sustrato	Producto	Especie*	Referencias
AD 9OH-AD 17β-hidroxi-5α-androstan-3-ona 5α-androstan-3,17-diona progesterona 11β-cortisol	ADD 9OH-ADD 17β-hidroxi-5α-1-androsten-3-ona 5α-1-androsten-3,17-diona 1,4-pregnadien-3,20-diona 11β,17α,20-trihidroxi-1,4-pregnadien-3,20-diona	*Rhodococcus. erythropolis SQ1*	Knol *et al.* 2008 van der Geize *et al.* 2002a
AD	ADD	*Rhodococcus opacus (Nocardia opaca)*	Groh, Komel *et al.* 1980
AD	ADD	*Rhodococcus equi (Nocardia restrictus)*	Abul-Hajj *et al.* 1978
AD 4-androsten-3,11,17-triona 4-pregnen-3,20-diona cortisona testosterona	ADD 1,4-androstadien-3,11,17-triona 1,4-pregnadien-3,20-diona 17α,21-dihidroxi-1,4-pregnadien-3,11,20-triona 17α-hidroxi-1,4-androstadien-3-ona	*Pimelobacter simplex (Corynebacterium simplex) (Arthrobacter simplex)*	Abul-Hajj *et al.* 1978 Choi, Molnar *et al.* 1995
AD 11β-hidroxi-4-androsten-3,17-diona 19-nor-4-androsten-3,17-diona androsterona 11β-hidroxitestosterona 11α-hidroxiprogesterona cortisona 11-desoxi-17α-hidroxicortisona corticosterona hidrocorticosterona	ADD 11β-hidroxi-1,4-androstadien-3,17-diona 19-nor-1,4-androstadien-3,17-diona 3α-hidroxi-1-androsten-17-ona 11β,17β-dihidroxi-11,4-androstadien-3-ona 11α-hidroxi-1,4-pregnadien-3,20-diona 17α,21-dihidroxi-1,4-pregnadien-3,11,20-triona 11-desoxi-17α,21-dihidroxi-1,4-pregnen-3,20-diona 17α,21-dihidroxi-1,4-pregnadien-3,20-diona 11β,17α,21-trihidroxi-1,4-pregnadien-3,20-diona	*Gordonio rubripertincto (Nocardia corallina)*	Itagaki, Wakabayashi *et al.* 1990
AD progesterona 16α,17α-epoxiprogesterona colesten-4-en-3-ona	ADD 1,4-pregnadien-3,20-diona	*Gordonio neofelifaecis* NRRL B-59395	Zhang, Ren *et al.* 2015
AD 11β-hidroxi-4-androsten-3,17-diona testosterona 11α-hidroxitestosterona 17α- hidroxiprogesterona 4-pregnen-3,20-diona desoxicorticosterona	ADD 11β-hidroxi-1,4-androstadien-3,17-diona 17α-hidroxi-1/4-androstadien-3-ona 11α,17β-dihidroxi-1,4-androstadien-3-ona 17α-hidroxi-1,4-pregnadien-3,20-diona 1,4-pregnadien-3,20-diona 21-hidroxi-1,4-pregnadien-3,20-diona	*Comamonas testosteroni*	Levy y Talalay 1955
AD 9OH-AD	ADD 9OH-ADD	*Mycobacterium neoaurum* NwIB	Wei, Fan *et al.* 2016
AD	ADD	*Corynebacterium crenatum* SYPA 5-5/pXMJ19-ksddII	Zhang, Wu *et al.* 2016
AD	ADD	*Ilyonectria destructans*	Abul-Hajj *et al.* 1978
AD	ADD	*Aspergillus fumigatus* CICC 40167	Chen, Wang *et al.* 2012
ácido 3-oxo-5β-colan-24-oico 5β-androstan-3,17-diona	5β-androsta-1-en-3,17-diona	*Clostridium paraputrificum*	Aries, Goddard *et al.* 1971

* Entre paréntesis se especifican los nombres de los microorganismos en el momento en el que se publicó el estudio.

Los residuos catalíticos de interacción con el sustrato están conservados excepto para el caso de la Ser^{468} de Δ4-(5α)-KSTD de *R. jostii* que se alinea con la Tyr^{318} de KstD1 de *R. erythropolis* (Dijkstra, van Oosterwijk *et al.* 2016). Se han identificado genes homólogos de *kstDs* en 100 especies microbianas, pero solo 24 especies tienen genes homólogos de Δ^4-(5α)-*kstd* (Kisiela, Skarka *et al.* 2012).

6.1.5. Aplicaciones biotecnológicas de las KstDs

Los productos de las 3-cetoesteroides deshidrogenasas se han convertido en sintones altamente comercializados en la producción de esteroides de uso farmacéutico, como por ejemplo, el 1,4-adrostadien-3,17-diona (ADD) (Dijkstra, van Oosterwijk *et al.* 2016). Actualmente, en *Mycobacterium*, se producen una variedad de sintones de esteroides farmacéuticos a través de la degradación de esteroides naturales (p. ej. colesterol y fitoesteroles) como una alternativa a la síntesis química (Galán, Uhía *et al.* 2016). La posibilidad de inactivar o sobreexpresar la actividad de las KstDs de un modo racional permite la acumulación selectiva de compuestos de interés (Wei, Wang *et al.* 2010, Galán, Uhía *et al.* 2016).

En la industria, la separación de AD del ADD es uno de los principales obstáculos para la transformación microbiana a partir de fitoesteroles. Con el objeto de aumentar la cantidad de ADD producido a partir de fitoesteroles en *M. neoaurum* NwIB-01 se ha sobreexpresado el gen de la KstD del propio organismo. La capacidad de producción de ADD aumentó con una relación molar de ADD en los productos del 98%, obteniendo de este modo cepas de biotransformación a partir de fitoesteroles con una alta pureza del producto (Wei, Wang *et al.* 2010, Wei, Fan *et al.* 2014). Igualmente, se han construido mutantes de *M. smegmatis* mc2155 mediante la deleción de los genes *MSMEG_6039* (*kshB1*$_{MSMEG}$) y *MSMEG_5941* (*kstD1* $_{MSMEG}$). A partir de la transformación de esteroles naturales (p. ej. fitosteroles), el doble mutante en *kshB1*$_{\text{MSMEG}}$ /*kstD1*$_{MSMEG}$ es capaz de producir y acumular AD, mientras que el mutante simple de *kshB1*$_{\text{MSMEG}}$ acumula ADD (Galán, Uhía *et al.* 2016). Asimismo, con el objeto de acumular eficientemente ADD a partir de AD, se creó una cepa recombinante de *Corynebacterium crenatum* (*C. crenatum* SYPA 5-5/pXMJ19-ksddII), que sobreexpresa la KstD (KSDD) de *M. neoaurum*, dando lugar un sistema de biotransformación más rápido y eficiente (Zhang, Wu *et al.*, 2016).

6.2. Las 3-cetoesteroide-9α-hidroxilasas (KshABs)

La 3-cetoesteroide-9α-hidroxilasa [Ksh; 3-ceto-4-eno esteroide: NADH: oxígeno oxidorreductasa (hidroxilante); EC 1.14.13.142] es una Rieske monooxigenasa (RO) clase IA. Las ROs catalizan una variedad de reacciones en las que uno o dos átomos de oxígeno se incorporan a un sustrato orgánico de una forma estero y regio específica (Barry y Challis 2013). Las monoxigenasas son sistemas de dos componentes que consisten en una oxigenasa terminal (KshA) y una ferrodoxin reductasa (KshB). KshA tiene un dominio de unión tipo Rieske [$Fe_2S_2Cys_2His_2$] y un hierro divalente no-hemo (Fe^{2+}), mientras que KshB es una flavoproteína que contiene un dominio de unión a NAD y un grupo hierro-azufre [$Fe_2S_2Cys_4$] (Capyk, D'Angelo *et al.* 2009, Capyk, Strynadka *et al.* 2010, Petrusma, van der Geize *et al.* 2014).

Las dos proteínas se unen mediante una cadena de transporte de electrones: NAD(P)H dona electrones a la flavina de la reductasa y estos electrones se transfieren a su vez a través de los grupos Fe-S a la oxigenasa, subunidad que lleva a cabo la hidroxilación (Batie, LaHaie *et al.* 1987, Ferraro, Gakhar *et al.* 2005, Tarasev y Ballou 2005).

KshAB es una enzima clave en el catabolismo bacteriano de esteroides puesto que se encarga del paso de AD a 9-OHAD o de ADD a 9OH-ADD, siendo este último un compuesto inestable que se hidroliza espontáneamente dando lugar a 3-HSA. Esta actividad -junto a la de KstD- lleva a la apertura del anillo B con aromatización simultánea del anillo A de la estructura esteroidea (Figura 8) (van der Geize, Hessels *et al.* 2002, Capyk, D'Angelo *et al.* 2009). Las 3-cetoesteroide-9α-hidroxilasas, al igual que las KstDs, se encuentran en actinobacterias degradadoras de esteroides tales como *R. erythropolis* (van der Geize, Hessels *et al.* 2002), *R. rhodochrous* (Petrusma, Dijkhuizen *et al.* 2009), *M. tuberculosis* (Capyk, D'Angelo *et al.* 2009) o *M. smegmatis* (Andor, Jekkel *et al.* 2006). Como suele ser característico en los genomas de estos géneros, se han encontrado múltiples copias de genes que codifican KshAs. Por ejemplo, hay 5 variantes en *R. rhodochrous* (Wilbrink, Petrusma *et al.* 2011) y *Mycobacterium sp* VKM Ac-1817D (Bragin, Shtratnikova *et al.* 2013); 3 en *R. erythropolis* (van der Geize, Hessels *et al.* 2008); y 4 en R. jostii (van der Geize, Yam *et al.* 2007, Mathieu, Mohn *et al.* 2010); aunque en *M. tuberculosis* solo se ha encontrado una única copia (Yang, Dubnau *et al.* 2007). La elaboración de árboles filogenéticos de estas enzimas ha ayudado a identificar varios grupos, como se puede ver en la Figura 9.

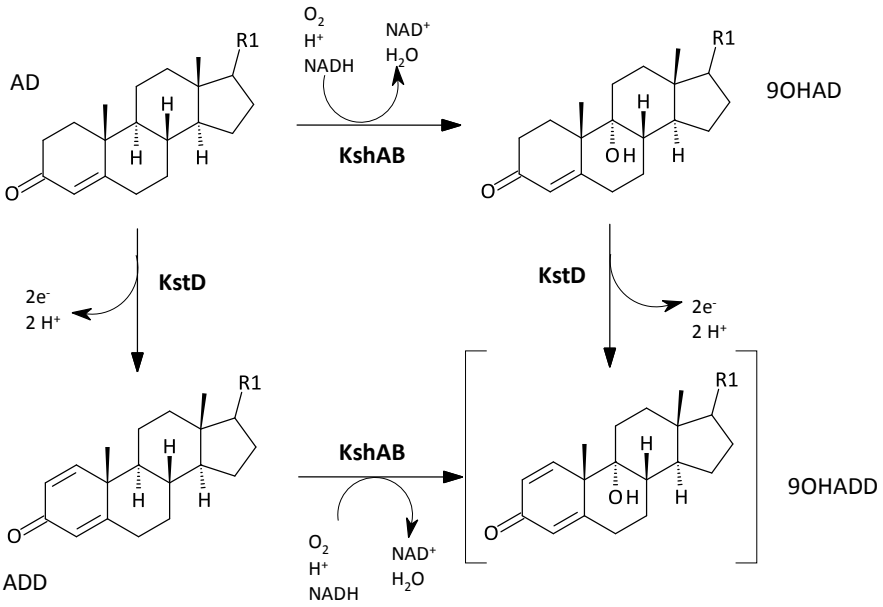


Figura 8. Reacciones catalizadas por las 3-cetoesteroide-9α-hidroxilasa (KshAB) y la 3-cetoesteroide-Δ^1-deshidrogenasas (KstD). Abreviaturas: AD: 4-androsten-3,17-diona; ADD: 1,4-androstadien-3,17-diona; 9OH-AD: 9α-hidroxi-4-adrosten-3,17-diona; 9OH-ADD: 9α-hidroxi-1,4-androstadien-3,17-diona; NADH: nicotin adenin dinucleótido.

Cada isoenzima parece tener una función distinta para diferentes pasos en la ruta catalítica de rhodococci. Así en *R. erythropolis* SQ1, KshA1 cataliza la conversión de AD en ADD, mientras que KshA2 parece estar involucrada en disminuir los niveles de ADD (van der Geize, Hessels *et al.* 2008). Igualmente, si se mutan *kshA* o *kshB* en *M. tuberculosis* éste deja de crecer en presencia de esteroides como única fuente de carbono y es rápidamente eliminado por los macrófagos (Hu, Van Der Geize *et al.* 2010). En *R. rhodochrous*, KshA1 y KshA5 están relacionadas con el metabolismo del ácido cólico, ya que el mutante sin actividad KshA (mutante de las 5 KshAs; *ΔkshA*), es incapaz de crecer en presencia de esteroides o ácido cólico como fuente de carbono. Sin embargo, el crecimiento en ácido cólico se recupera al complementar con KshA1 o KshA5. Por otro lado, las KshA2, KshA3 y KshA5 están relacionadas con el metabolismo del colesterol, dado que son capaces de complementar y restablecer el crecimiento en este compuesto. Cabe destacar que solo KshA2 y KshA5 pueden recuperar el crecimiento del mutante *ΔkshA* en AD y progesterona (Petrusma, Hessels *et al.* 2011). Estos datos ponen de manifiesto el hecho de que estas enzimas no solo están relacionadas con tipos específicos de esteroles sino también con pasos determinados de la ruta catabólica.

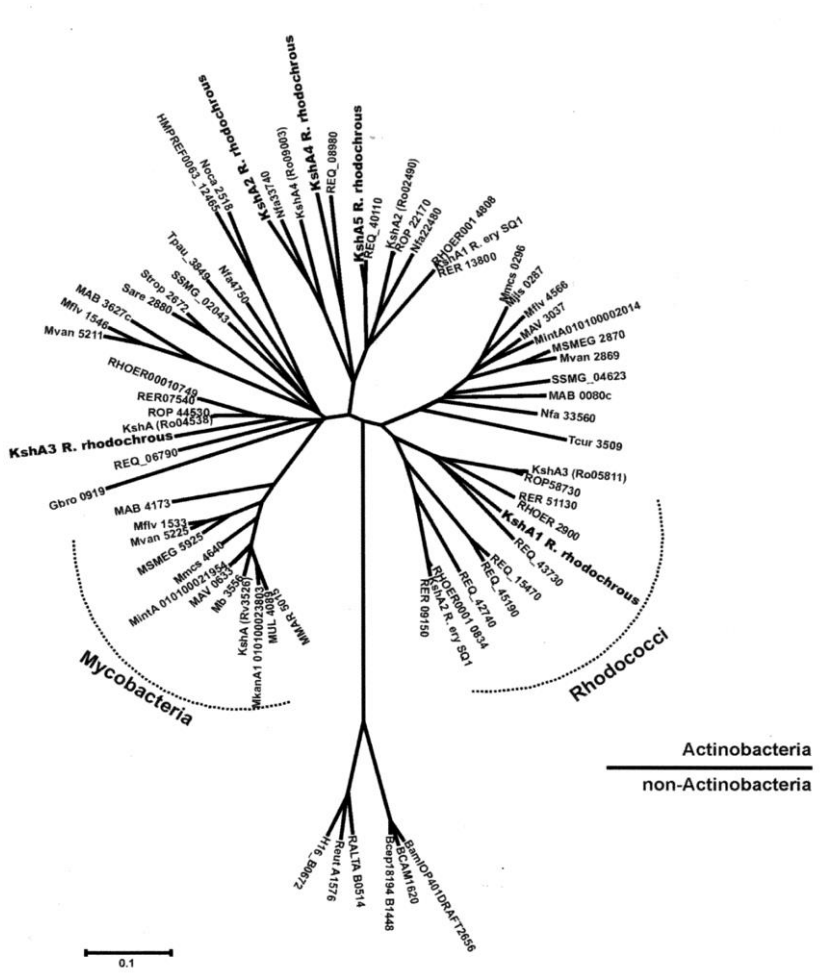


Figura 9. Árbol filogenético de las enzimas KshAs de *R. rhodochrous* y sus homólogos en otras actinobacterias y especies de Burkholderia, publicado por Petrusma *et al.* (2011).

6.2.1. Regulación de las KshABs

La regulación de las KshAB ha sido escasamente estudiada. Sin embargo se ha determinado su inducción en algunos organismos, como es el caso de *R. rhodochrous* DSM 43269 en el que cada isoenzima se induce por un tipo distinto de compuesto. Así KshA1 se induce por ácido cólico y algo menos por colesterol; la KshA3 solo se induce en presencia de AD; la KshA4 en presencia de ácido cólico y colesterol, pero en AD su inducción es menor. Finalmente, KshA5 se induce por AD o progesterona y su inducción se ve reducida por ácido cólico y colesterol (Petrusma, Hessels *et al.* 2011). En el caso de *R. jostii* RHA1, las cuatro KshAs identificadas se inducen por el 7-cetocolesterol y colesterol, pero además las KshA3 y KshA4 también se inducen en presencia de colato (Mathieu, Mohn *et al.* 2010). Todo esto apunta a que cada isoenzima está finamente regulada y relacionada con una determinada función.

6.2.2. Estructura de las KshAs

La estructura de KshA de *M. tuberculosis* ha sido elucidada por cristalografía. Esta contiene un grupo Rieske hierro-azufre (2Fe-2S) N-terminal seguido de un dominio catalítico más grande que incluye un hierro no-hemo divalente (Fe^{2+}) localizado en el centro del sitio activo y que está relacionado con la unión y activación del oxígeno y la hidrolización del sustrato (Figura 10). La enzima funcional está organizada en trímeros que forman un plegamiento α3 típico. La interacción de las subunidades es cabeza-pie donde el grupo Rieske hierro-azufre está en proximidad al hierro (Fe^{2+}) del centro activo. El dominio Rieske (residuos 24-153) consiste en tres hojas β anti paralelas y es precisamente la tercera la que contiene el dominio Rieske. Un ion ferroso del dominio está coordinado por la Cys^{67} y Cys^{86}, el otro está coordinado por la His^{69} e His^{89} (Capyk, D'Angelo *et al.* 2009).

El domino catalítico (residuos 154-374) consiste en ocho hojas β antiparalelas, siendo flanqueadas por una serie de α-hélices en la cara distal del dominio Rieske. Este dominio se distingue de otras ROs por su porción C-terminal (338-374) que tiene una hélice terminal más larga situada en la superficie externa del trímero. Esta estructura se localiza junto a los residuos 103-106 del dominio Rieske del protómero adyacente, lo que parece estabilizar al trímero. La orientación del acceso al canal del sitio activo y la posición de la boca del canal son distintas a otras ROs. El hierro mononuclear del sitio activo se encuentra en el centro y el solvente solo accede a él a través de un canal conformado básicamente de residuos hidrofóbicos. El ion hierro está coordinado por la His^{181}, His^{186} y el Asp^{304}. En las KshAs el sitio de unión al sustrato se encuentra muy elongado en comparación a otras ROs. Esta elongación es consistente con el tamaño del sustrato (p. ej. AD o ADD) que es grande en comparación a los sustratos conocidos de otras ROs. Los residuos identificados de interacción con el sustrato son: Val^{176}, Gln^{204}, Tyr^{232}, Met^{238}, Asn^{240}, Asn^{257}, Phe^{301} y Trp^{308}. Todos ellos se encuentran altamente conservados en las ROs que se han anotado como candidatas a enzimas catalizadoras de esteroides (Capyk, D'Angelo *et al.* 2009).

El mecanismo catalítico de las KshABs todavía no se ha esclarecido, aunque varios estudios de las ROs ofrecen una cierta idea de los pasos implicados. El primero es la unión del sustrato al sitio activo y la reducción del hierro no-hemo. Después de la unión al sustrato, el grupo Rieske hierro-azufre también debe estar en un estado reducido. Únicamente cuando la oxigenasa está en este estado, el oxígeno puede unirse al centro metálico. Esta disposición evita el desacoplamiento ya que todos los componentes de la reacción son necesarios para la activación de oxígeno. El ciclo catalítico necesita de la entrada de dos equivalentes reducidos que son

transferidos a través de una red que conecta el grupo Fe_2S_2 de la KshB, el dominio Rieske de KshA y el hierro catalítico de KshA. Se piensa que el Asp^{178} hace de elemento clave en la transferencia de electrones entre los metalocentros del grupo Rieske y el hierro divalente mononuclear de la KshA (Capyk, Strynadka *et al.* 2010, Petrusma, van der Geize *et al.* 2014). Este aspartato también se ha relacionado con el mantenimiento del estado de protonación y el potencial de reducción del centro Rieske (Capyk, Strynadka *et al.* 2010).

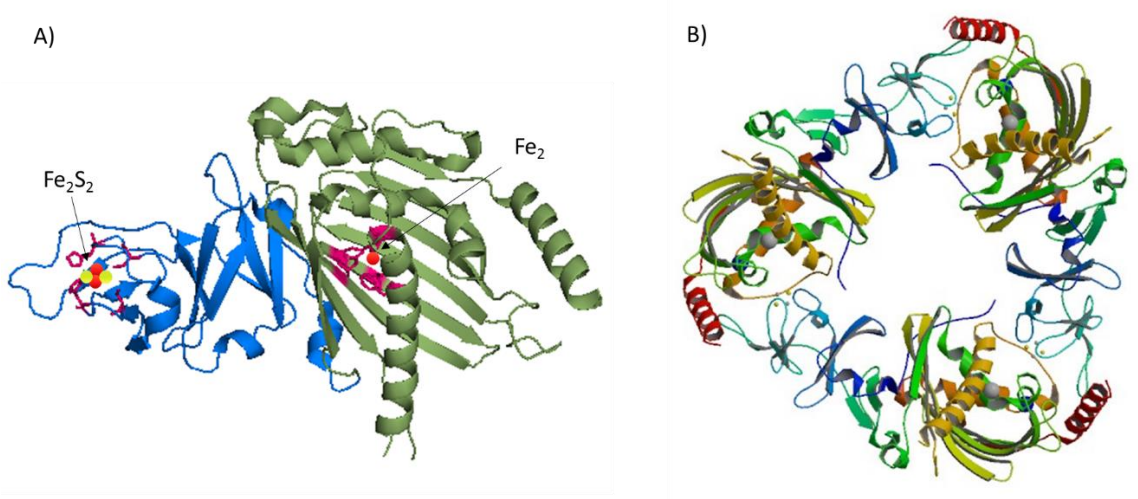


Figura 10. Modelo a partir del cristal de KshA de *M. tuberculosis* (PDB: 2ZYL). **A)** Monómero; el dominio Rieske está representado en azul y el dominio catalítico en verde, los residuos de unión al grupo hemo del dominio Rieske y al ion hierro divalente no hemo del dominio catalítico están resaltados en rojo; **B)** trímero visto desde un eje simétrico, figura de ensamblaje biológico tomada del PDB (Capyk, D'Angelo *et al.* 2009).

6.2.3. Tipos KshAs y sustratos

La mayor parte de los estudios realizados con las KshAs se han llevado a cabo en *Rhodococcus* o *Mycobacterium*. Los análisis para determinar la especificidad de sustrato en *R. rhodochrous* DSM43269 de KshA1 y KshA5 han demostrado que KshA1 tiene mayor especificidad por esteroides con cadena lateral en C17 (p. ej. 3-oxo-23,24-bisnorcolest-1,4-dien-22-oato; 1,4-BNC), mientras que su homólogo KshA5 presenta la especificidad más alta por sustratos sin cadena lateral (p. ej. AD y testosterona) y su acción es inhibida por sustratos como el ADD o el 4-BNC.

Tabla 4. Lista de reacciones enzimáticas catalizadas por KshAB en diferentes organismos (última revisión, 29 de febrero del 2017).

Sustrato	Producto	Especie	Referencias
AD	9OH-AD	*Rhodococcus rhodochrous* DSM 43277	Petrusma *et al.* 2009 Petrusma *et al.* 2011 Petrusma *et al.* 2012
ADD	9OH-ADD		
11β-hidrocortisona	9α-hidroxi-11β-hidrocortisona		
5α-androsten-3,17-diona			
1-(5α)-androsten-3,17-diona	9α-hidroxi-5α-androsten-1-en-3,17-diona		
19-nor-4-androsten-3,17-diona	9α-hidroxi-19-nor-4-androsten-3,17-diona		
progesterona			
testosterona			
5β-androsten-3,17-diona	9α-hidroxi-5β-androstan-3,17-diona		
5α-androstan-3,17-diona	9α-hidroxi-5α-androstan-3,17-diona		
5α-androstan-17β-ol-3-one	9α,17β-dihidroxi-5α-androstan-3-ona		
4-pregnen-3,20-diona	9α-hidroxi-4-pregnen-3,20-diona		
4-colesten-3-ona			
4-androsten-17β-ol-3-one	9α,17β-dihidroxi-4-androsten-3-ona		
ácido 3-oxo-23,24-bisnocolestan-4-ene-22-oico	ácido 3-oxo-9α-hidroxi-23,24-bisnocolestan-4-ene-22-oico		
ácido 23,24-bisnocolestan-4-ene-22-oico			
ADD	9OH-ADD	*Rhodococcus erythropolis* SQ2	van der Geize *et al.* 2003 van der Geize *et al.* 2008
AD	9OH-AD		
AD	9OH-AD	*Mycobacterium tuberculosis* H37Rv	Capy *et al.* 2009 Capy *et al.* 2011
ADD	9OH-ADD		
3-oxo-23,24-bisnocolestan-4-ene-22-oil-CoA	3-oxo-9α-hidroxi-23,24-bisnorcolestan-4-ene-22-oil-CoA		
3-oxo-23,24-bisnocolestan-4-ene-22-oil-CoA	3-oxo-9α-hidroxi-23,24-bisnorcolestan-4-ene-22-oil-CoA		
3-oxo-23,24-bisnocolestan-4-ene-22-oato	3-oxo-9α-hidroxi-23,24-bisnorcolestan-4-ene-22-oato		
3-oxo-23,24-bisnocolestan-1,4-dien-22-oato-CoA	9α-hidroxi-3-oxo-23,24-bisnorcolestan-1,4-dien-22-oato-CoA		
3-oxo-23,24-bisnocolestan-1,4-dien-22-oato	9α-hidroxi-3-oxo-23,24-bisnorcolestan-1,4-dien-22-oato		
ácido 3-oxo-23,24-bisnorcola-1,4-dien-22-oico-CoA			
ácido 3-oxo-23,24-bisnorcola-1,4-dien-22-oico			
ácido 3-oxo-23,24-bisnorcol-4-en-22-oico-CoA			
ácido 3-oxo-23,24-bisnorcol-4-en-22-oico			
ADD	9OH-ADD	*Mycobacterium neoaurum* NwIB	Wei *et al. 2010*
AD	9OH-AD		

Por otra parte, los sustratos fisiológicos de las KshAB en *M. tuberculosis* son intermediarios (CoA-tioesteres) de la degradación de la cadena lateral del colesterol; lo que demuestra que la degradación ocurre simultáneamente a la ruptura del anillo (Capyk, Casabon *et al.* 2011, Penfield, Worrall *et al.* 2014). La estructura cristalina de estas enzimas unidas al sustrato muestra bolsillos de unión a esteroides muy conservados, en los cuales el C17 de los sustratos está orientado hacia la apertura del sito activo. Sin embargo, el hierro catalítico y las α-hélices que contienen los ligandos de la KshA5 se encuentran desplazados en el complejo sustrato-KshA5 en comparación con la KshA5 libre de sustrato, lo que sugiere que las Rieske oxigenasas pueden tener una naturaleza dinámica similar al citocromo P450 (Penfield, Worrall *et al.* 2014). En la Tabla 4 se recogen las 3-cetoesteroide-9α-hidroxilasa descritas hasta la fecha con su sustrato y -en los que se conoce- su producto.

6.2.4. Aplicaciones biotecnológicas de las KshABs

Los compuestos esteroideos hidroxilados tienen una gran relevancia tanto en el sector industrial como en el campo de la medicina debido a que son altamente bioactivos. La hidroxilación de esteroides en C-9α junto con otra como C-7α o C-11β son de gran interés para la industria farmacéutica (Donova y Egorova 2012) por ser compuestos claves para la producción de medicamentos (p. ej. corticoides, agentes antitumorales, antiinflamatorios, antimicrobianos). Todo esto hace que la producción de sintones hidroxilados (p. ej. 9OH-AD) a partir de esteroides como los fitoesteroides confiera a la bioconversión microbiana un alto interés comercial. Por ejemplo, con el objetivo de producir 9OH-AD a partir de fitoesteroles se creó una cepa de *M. neoaurum* ATCC 25795 mutante en los genes *kstDs.* El 9OH-AD producido estaba altamente mezclado con AD (un subproducto de la bioconversión). Con el objetivo de aumentar la pureza del producto se sobreexpresó el gen de la *kshA* lo que mejoró la producción de este sintón en un 95% (Yao, Xu *et al.* 2014). Otra aplicación biotecnológica es la coexpresión de KshA de *M. neoaurum* JC-12 y de la glucosa 1-dehydrogenase (GDH) para construir un sistema de regeneración de NADH en *Bacillus subtilis* 168 donde se ha obtenido una cepa de bioconversión eficiente de AD en 9OH-AD (Zhang, Rao *et al.* 2016).

IV. OBJETIVOS

OBJETIVOS

En el laboratorio se aisló y caracterizó, previamente, la cepa *Rhodococcus ruber* Chol-4 a partir de lodos de una depuradora de aguas en Ciudad Real, España. Esta cepa es capaz de crecer en presencia de una amplia gama de esteroides como única fuente de carbono y energía (Fernández de las Heras, García Fernández *et al.* 2009). Los datos preliminares de secuenciación del genoma de cepa *Rhodococcus ruber* Chol-4 permitieron la identificación de varios genes involucrados en la ruta de degradación del colesterol y otros esteroides, entre ellos una colesterol oxidasa (*choG*) y tres 3-cetoesteroide-Δ^1-deshidrogenasas (*kstDs*). El estudio del gen *choG* demostro que codifica una colesterol oxidasa funcional extracelular involucrada en la conversión de colesterol en colestenona. Por otro lado, se caracterizaron bioquímicamente tres KstDs identificadas mediante la generación de mutantes e detección de intermediarios. Estos datos demostraron que KstD1 y KstD2 son las responsables de la conversión de AD en ADD y de 9OHAD en 9OHADD, también que KstD3 puede estár relacionada con la conversión del colesterol en ADD por una ruta distinta a la de KstD1 y KstD2 en *R. ruber* (Fernández de las Heras, van der Geize *et al.* 2012).

Con estos antecedentes, el objetivo principal del presente trabajo consiste en la profundización del estudio de genes implicados en la ruta de degradación de esteroides en *Rhodococcus ruber* Chol-4. Para su consecución, se marcaron los siguientes objetivos parciales (entre paréntesis se especifican los Capítulos, dentro del apartado de Resultados, en los que se ha abordado cada objetivo):

1. Estudio del genoma de *Rhodococcus ruber* Chol-4.

Con el objeto de ampliar nuestro conocimiento sobre la cepa de estudio y potenciar con ellos sus posibilidades biotecnológicas, se realizó un análisis más exhaustivo del genoma de *Rhodococcus ruber* Chol-4. Los objetivos parciales para llevar a cabo este trabajo son (Capítulo 1):

- **1.1.** Notación, análisis y caracterización del genoma.
- **1.2.** Identificación los genes relacionados con diferentes rutas catabólicas.
- **1.3.** Análisis de las capacidades de crecimiento.

2. Identificación, aislamiento y caracterización de genes implicados en la ruta de degradación del anillo esteroideo de *Rhodococcus ruber* Chol-4.

Centrándonos concretamente en las enzimas que participan en el proceso de ruptura del anillo

esteroideo, se prestará especial atención a los genes codificantes de las actividades 3-cetoesteroide-Δ^1-deshidrogenasa (KstD) y 3-cestoesteroide-9α-hidroxilasa (KshAB). Para ello se establecieron los siguientes objetivos parciales:

2.1. Estudio de KstD en *Rhodococcus ruber* Chol-4 (Capítulo 2).

2.1.1. Expresión heteróloga de las tres KstD de *Rhodococcus ruber* Chol-4 en *Rhodococcus erythropolis* 3014.

2.1.2. Caracterización de la actividad de las tres KstDs en presencia de diferentes sustratos esteroideos.

2.1.3. Identificación *in silico* de los promotores putativos de las tres enzimas KstDs y su caracterización funcional.

2.2. Estudio de la funcionalidad de las 3-cetoesteroide-9α-hidroxilasas (KshAB) de *Rhodococcus ruber* Chol-4 (Capítulo 3).

2.2.1. Identificación en el genoma y generación de mutantes defectivos en la actividad de las enzimas KshAB.

2.2.2. Análisis de la capacidad de crecimiento de las cepas mutantes generadas en presencia de diferentes esteroides como única fuente de carbono y energía.

2.2.3. Detección de posibles intermediarios acumulados como consecuencia de las mutaciones generadas.

3. Aplicaciones biotecnológicas.

Con el fin de aplicar los estudios derivados de los objetivos anteriores en la mejora de los procesos de biotransformación para la obtención de esteroides de interés industrial en el género *Rhodococcus*, se ha planteado la expresión de los genes *kstD* y caracterización funcional de KstDs de otros orígenes para la generación de una cepa de *Rhodococcus* hiperdeshidrogenadora (Capítulo 4).

V. MATERIALES Y MÉTODOS

MATERIALS AND METHODS

1. Materials and Methods Chapter 1: New insights into *Rhodococcus rube*r Chol-4 genome

1.1. DNA Manipulation and sequencing

Chromosomal DNA extraction from *R. ruber* strain Chol-4 was performed using the CTAB procedure (Fernandez de las Heras, et al 2012). Manipulation of genomic DNA was carried out according to standard protocols (Sambrook *et al.*, 1989), and the extracted DNA was purified three times to achieve highest purity and quality for subsequent sequencing of the complete genome. The pyrosequencing of the genomic DNA was done by BGI - Hong Kong Laboratory (Hong Kong, China) using Solexa/Illumina. The program SPAdes (Bankevich, Nurk *et al.* 2012) was employed to assemble the reads, and manual assembly was carried out in order to reduce the size of contigs.

1.2. Bacterial strains and culture conditions

The bacterial strains and plasmids used in this work are listed in Table 1. *Escherichia coli* cells were grown at 37 °C in Luria Bertani (LB) at 250 rpm (Sambrook, Fritsch *et al.* 1989). *Rhodococcus ruber* and its derived mutant strains were routinely grown in LB or minimal medium (Medium 457 of the DSMZ, Braunschweig, Germany) containing the desired carbon and energy source under aerobic conditions at 30 °C in a rotary shaker (250 rpm) for 1-3 days. Where appropriate antibiotic were added: ampicillin (100 µg/mL), nalidixic acid (15 µg/mL) or kanamycin (25-50 µg/mL for *E. coli* or 200 µg/mL for *Rhodococcus*). For the growth experiments, a LB pre-grown culture was washed two times with minimal medium prior to inoculation of 10 mL of fresh minimal medium (initial DO_{600nm}= 0.05) supplemented with an organic compound as only source of energy and carbon. Volatile compounds such as indane, tetralin, isopropanol, 1,3-butanediol, 2,3-butanediol, xylene, benzene, ethylbenzene, toluene, phenylacetic acid or styrene were given as vapor. DHEA (dehydroepiandrosterone) and cholic acid were previously dissolved in 16.5 mM methyl-β-cyclodextrin (CD) to form inclusion complexes following a modification of a previously reported method (Klein, Gimpl *et al.* 1995) and prepared as described (Fernández de las Heras, Mascaraque *et al.* 2011).

Competent and electrocompetents cells of *E. coli* were prepared and transformed as previously described (Sambrook, Fritsch *et al.* 1989). +Selection of transformed cells was carried out in LB plates supplemented with appropriate antibiotics.

Table 1. Bacterial strains and plasmids used in this work.

Strain and plasmids	Description	Reference
***E. coli* DH5α**	F' *endA1 hsdR17* (r_{K^-} m_{K^+}) *glnV44 thi-1 recA1 gyrA* (Nal^R) *relA1* Δ(*lacIZYA-argF*) *U169 deoR* (ϕ80*dlac*Δ(*lacZ*)*M15*).	Laboratory collection
***E. coli* GM48**	F^- THR LEU THI LACY GALK GALT ARA FHUA TSX DAM DCM GLNV44	Laboratory collection
***E. coli* S17.1**	*recA pro hsdR RP4-2-Tc::Mu-Km::Tn7.*	(Simon, Priefer *et al.* 1983)
pGem-T Easy	Ap^R	Promega
pK18mobsacB	Km^R, RP4mob,mobilizable cloning vector with *sacB* gene for positive selection	**(**Schafer, Tauch *et al.* 1994)
pK18A	pK18mobsacB with fragment A from the *nar cluster* cloned between *Eco*RI and *Xba*I.	This work
pK18AB	pK18mobsacB with fragments A and B from the *cluster nar* cloned between *Eco*RI and *Hind*III.	This work
pK18C	pK18mobsacB with the fragment C from the *pca cluster* cloned between *Eco*RI and *Xba*I.	This work
pK18CD	pK18mobsacB with fragments C and D from the *pca cluster* cloned between *Eco*RI and *Hind*III.	This work
***R. ruber* Chol-4**	Wild type phenotype, Nal^R	(Fernández de las Heras, García Fernández *et al.* 2009)
Nar R.ruber	Mutant in *nar* genes, Nal^R.	This work
PCA R. ruber	Mutant in *pca* genes, Nal^R.	This work

1.3. Mutagenesis of *R. ruber* strain Chol-4

Primers and conditions applied for this experiment are summarized in Table 2. Unmarked gene deletions were carried out as described previously in *R. erythropolis* SQ1 involving conjugative transfer of a mutagenic plasmid carrying the *sacB* selection system (van der Geize, Hessels *et al.* 2001). Specific set of primers were designed up- and downstream of each ORF in order to get PCR amplicons from isolated *R. ruber* strain Chol-4 genomic DNA with the ORF flanking regions containing the restriction sites *Eco*RI-*Xba*I (up) and *Xba*I-*Hin*dIII (down).

These amplicons were cloned into pGEM-T-Easy vectors and both combined in order to get a fragment *Eco*RI-*Hin*dIII containing a truncated ORF and its flanking regions, but keeping the reading frame to avoid polar effects. It was necessary to transform in *E.coli* GM48 in order to avoid Dam methylation in the *Xba*I site. *Eco*RI-*Hind*III inserts were moved to pK18mobsacB plasmid (Schafer, Tauch *et al.* 1994) to construct the mutagenic plasmids pK18(U+D), used for the deletion of the ORF from *R. ruber* strain Chol-4 chromosome.

The mutagenic plasmids were introduced into *E. coli* S17.1 and mobilized to *R. ruber* strain Chol-4 by conjugation as previously described (Fernández de las Heras, van der Geize *et al.* 2012). *R. ruber* transconjugants that have integrated the plasmid by homologous recombination were selected. The ORF deletion was achieved as a result of a second spontaneous homologous recombination process within the genome of *R. ruber* strain Chol-4. A colony PCR detection was performed afterwards in order to check this deletion in the *nar* and *pca* mutant strains.

Table 2. Primers used in this work

Primer	Sequence	PCR conditions	Use
CH407	AGATCTCCCGGGGAATTCGGTCGAGCTCGCCCGGTTCGAC	Tm 56 °C, 1.5 min, 30 cycles.	Amplification of fragment A (*pca genes)*
CH408	AGTACTGATATCTCTAGATCGTTCTCGAAGGGGGTGGCCT		
CH409	AGTACTGATATCTCTAGACCCAATACGGCGGAGCAGCAGGAA	Tm 60 °C, 1.5 min, 30 cycles.	Amplification of fragment B (*pca genes)*
CH410	GCTAGCCTGCAGAAGCTTATCCTGGAACGGCCGGGTGAACTC		
CH411	AGATCTCCCGGGGAATTCGTCGACCGCGGCGATGTCCACC	Tm 56 °C, 1.5 min, 30 cycles.	Amplification of fragment C (*nar genes)*
CH412	AGTACTGATATCTCTAGACTCTACGTGCAACAGGGCGCAA		
CH413	AGTACTGATATCTCTAGACTGCTCCGGCGTGAGCCGACG	Tm 56 °C, 1.5 min, 30 cycles.	Amplification of fragment D (*nar genes)*
CH414	GCTAGCCTGCAGAAGCTTCCGTCGTGCGCGAGGGGCTTTC		
CH438	TGTACTTCCCCGACGACCCGAT	Tm 58 °C, 1.5 min, 30 cycles	Checking *pca* cluster deletion
CH439	TCCTCGCTCACCTCGCCGACTT		

1.4. Genome analysis and annotation

Homology searches were performed using the BLAST server (http://blast.ncbi.nlm.nih.gov/Blast.cgi) at GenBank. The annotation of the genome was carried out using the GenBank tool PGAP and on-line service RAST (http://rast.nmpdr.org/). The complete genome sequence has been deposited at GenBank under accession number NZ_ANGC00000000.2. The program Circos was used to visualize genomic data (Krzywinski, Schein *et al.* 2009).

1.5. Pulsed Field gel electrophoresis

PFGE was performed from 10 mL cells grown at OD_{600nm} of 0.8-1.0, centrifuged and suspended in 0.5 mL of cell suspension solution (10 mM Tris-HCl pH 7.2, 20 mM NaCl, 100 mM EDTA). Plugs containing the cells were prepared with 1.5% agarose and placed in lysis buffer (1 mg/mL lysozyme, 10 mM Tris-HCl pH 7.2, 50 mM NaCl, 100 mM EDTA, 0.2% DOC, 0.5% N-laurylsarcosine sodium salt, 0.06 g/L RNase) and incubated for 1 h at 37 °C with soft shaking. Lysis was followed by two washes in 20 mM Tris-HCl pH 8 and 50 mM EDTA. The plugs were placed in 3 mL proteinase solution (1 mg/mL proteinase K, 100 mM EDTA pH 8.0, 1% N-lauryilsarcosine sodium salt, 0.2% DOC) and were incubated with gently shaken at 42 °C for 18 hours. After removing the proteinase solution, 9 mL TE containing 40 µg/mL PMSF were added and kept at 50 °C for one hour, repeating the whole process two times. After washing twice for 15 minutes in 20 mM Tris-HCl pH 8, 50 mM EDTA. Plugs were resolved by PFGE on a contour-clamped homogeneous electric field (CHEF) II Mapper system (Bio-Rad Laboratories) in 0.5× Tris-borate-EDTA (TBE) and

the following running conditions: 6 V/cm for 18-24 h at 13 °C, with a 50-s switch time. Gels were stained in Gel Red solution (5 min) and photographed under UV light.

1.6. Phytosterol consumption followed by MS-HPLC

R. ruber was grown at 30 °C and 200 rpm, in 25 mL of minimal medium (M457 of the DSMZ, Braunschweig, Germany) supplemented with a mixture of industrial phytosterols in powder (around 0.7 mg/mL), kindly given by Gadea S.A. Samples at different times were collected until OD_{600nm} was of 1.9, 2 mL aliquots were taken and 1mg of pregnenolone was added as internal control of the extraction. The steroid fraction was extracted twice with 2 mL of chloroform. HPLC and MS determination was carried out in the Chromatography Service of "Centro de Investigaciones Biológicas "(CIB). The relative peak area was calculated as the ratio between the HPLC peak obtained for each phytosterol (brassicasterol, campesterol, stigmasterol and β-sitosterol) and the peak of pregnenolone used as internal control.

2. Materials and Methods Chapter 2: Functional differentiation of 3-ketosteroid Δ1-dehydrogenase isozymes in Rhodococcus ruber strain Chol-4

2.1. Bacterial strains, plasmids and growth conditions

The strains were routinely grown as previously described in section 1.2. For steroids growth experiments, a LB pre-grown culture was washed two times with minimal medium prior to inoculation. Cholesterol or AD (Sigma) were added directly to the minimal medium culture for growing and/or induction at 0.6 g/L and 0.44 g/L, respectively. Alternatively, Cholesterol and AD were previously dissolved in methyl-β-cyclodextrin (CD) (Klein, Gimpl *et al.* 1995) and prepared as described (Fernández de las Heras, Mascaraque *et al.* 2011).

For the promoter growth experiments, cells were plated in minimal medium M457 plates containing the desired carbon source and incubated at 30 °C for 3 days. Plasmids and bacterial strains used are listed in Table 3. Competent cells of *E. coli* DH5αF' and BL21 (DE3) were prepared as indicated in section 1.2.

Table 3. Bacterial strains and plasmids used in this work

Bacteria and plasmids	Description	Reference
Rhodococcus ruber strain Chol-4	Wild type phenotype, Nal^r	CECT7469
Rodococcus erythropolis strain *3014*	Wild type phenotype, Nal^r	CECT3014
E. coli DH5α	F' *endA1 hsdR17* ($r_K^- m_K^+$) *glnV44 thi-1 recA1 gyrA*(Nal^r) *relA1* Δ(*lacIZYA-argF*) *U169deoR* (ϕ80*dlac*Δ(*lacZ*)*M15*)	Laboratory collection
pBluescript II KS +	Cloning Vector, Ap^r	Stratagene
pGEM-T Easy Vector	Cloning vector, Ap^r	Promega
pTip-QC1	Expression *E. coli-Rhodococcus* shuttle vector, Ap^r, P_{TIPA} Chl^rREPAB (pRE2895)	(Nakashima and Tamura 2004)
pTip-*kstD1*	Expression plasmid harbouring *kstD1* (*Nde*I-*Bgl*II, 1.5kb)	This work
pTip-*kstD2*	Expression plasmid harbouring *kstD2* (*Nde*I-*Bgl*II, 1.6kb)	This work
pTip-*kstD3*	Expression plasmid harbouring *kstD3* (*Nde*I-*Bam*HI, 1.7kb)	This work
pNV119	*Nocardia-E. coli* replicative shuttle vector, Km^r	(Chiba, Hoshino *et al.* 2007)
pSEVA351	Cloning vector, Cm^r	(Silva-Rocha, Martinez-Garcia *et al.* 2013)
pNVS	*Nocardia-E. coli* replicative shuttle vector containing Teminator 1-mcs-terminator 0 from pSEVA351, Km^r (*Nhe*I-*Pci*I, 0.4Kb)	This work
pIJ773	Cloning vector carrying the Am^r gene	(Gust, Challis *et al.* 2003)
pNVSP1	pNVS containing *kstD1* promoter region and the sequence of the first seven amino acids in the mcs. (*Xba*I-*Pst*I, 0.4Kb)	This work
pNVSP2	pNVS containing *kstD2*promoter region and the sequence of the first seven amino acids in the mcs. (*Xba*I-*Pst*I, 0.2Kb)	This work
pNVSP3	pNVS containing *kstD3* promoter region and the sequence of the first seven amino acids in the mcs. (*Xba*I-*Pst*I, 0.1Kb)	This work
$pNVSP3^b$	pNVS containing *kstD3* minimal promoter region and the sequence of the first seven amino acids in the mcs. (*Kpn*I-*Pst*I, 0.05Kb)	This work
pNVSP1-A	pNVSP1 in frame to Am^r gene (*Nru*I-*Hind*III, 0.8Kb)	This work
pNVSP2-A	pNVSP2 in frame to Am^r gene (*Nru*I-*Hind*III, 0.8Kb)	This work
pNVSP3-A	pNVSP3 in frame to Am^r gene (*Nru*I-*Hind*III, 0.8Kb)	This work
$pNVSP3^b$-A	$pNVSP3^b$ in frame to Am^r gene (*Nru*I-*Hind*III, 0.8Kb)	This work
pNVSA	A derivative of pNVSP1-A with the promoter P1 region deleted	This work

2.2. Cloning of *kstD1*, *kstD2* and *kstD3* of *R. ruber* strain Chol-4 and heterologous expression in *Rhodococcus erythropolis* CECT3014 cells

The *kstD* ORFs were previously identified using the Bioedit program (Fernández de las Heras, García Fernández *et al.* 2009) and they were PCR amplified, from start to stop codon, using primers from Table 4. PCR was performed under standard conditions using High Fidelity PCR Enzyme Mix (Fermentas) with a specific high GC buffer (Roche) at 30 cycles of 1 min at 95 °C, 1 min at the desired Tm and 0.5-3 min at 72 °C (unless stated otherwise).

Table 4. Primers used in this work. Restriction sites are marked in bold.

Primer	Sequence	PCR conditions	Use
3exF 3exR	AA**CATATG**GTGGATTGGGCAGAGGAAT **AGATCT**GGGCGGCCGGCGTCGTCCTTAC	Tm 60 °C, 2 min, 30 cycles	*kstD1 Nde*I-*Bgl*II amplification for cloning and expression, 1.5 kb
2exF 2exR	AA**CATATG**GCGACCAATCCCGTACCGG AA**AGATCT**TCAGCGGGACTTCGCGGCGTCC	Tm 65 °C, 2 min, 30 cycles	*kstD2 Nde*I-*Bgl*II amplification for cloning and expression, 1.6 kb
1exF 1exR	AA**CATATG**ATCAAGCAGGAATACGACA AA**GGATCC**TCAGTGCTGTTCGACGGTCTCG	Tm 55 °C, 2 min, 30 cycles	*kstD3 Nde*I-*Bam*HI amplification for cloning and expression, 1.7 kb
CH112 CH113	CATCGACGACGCCCTGACCTACTA GAGCGTGTCGGCGGTCTTCC	Tm 55 °C, 1 min, 25 cycles	*kstD1* amplification in RT-PCR, 900 bp
CH106 CH107	CCAAGGGCTATTCCGACTACCA GAAAAGCCCGCCGCCGAAGATGTA	Tm 55 °C, 1 min, 25 cycles	*kstD2* amplification in RT-PCR, 800 bp
CH197 CH198	GTGCGCATCTGCTCGTTGTGTT GCCGGCAAGATCGACACCTACAT	Tm 55 °C, 1 min, 25 cycles	*kstD3* amplification in RT-PCR, 600 bp
CH190 CH191	CGACGGCAGCTGTACGAGACCT AACTGCCCGGGACGACCTTG	Tm 60 °C, 1 min, 25 cycles	ORF 4 amplification in RT-PCR, 1000 bases
CH192 CH193	GCTCGACACCACACCGCTGAA GCTCTCGCCGCGGTGGTATT	Tm 60 °C, 1 min, 25 cycles	ORF 5 amplification in RT-PCR, 1000 bases
CH344 CH345	ACCGGAAGCTGCCGGCCCT TGCCCCGTCGAGCACTGACC	Tm 58 °C, 0.5 min, 25 cycles	*Cyp450-kstD1* coexpression (RT-PCR) 242bp
CH505 CH506	GACATCGAGTTCACGGCCTA GGAGCCTTGCCGAAGTAGTC	Tm 60 °C, 1min, 30 cycles	*kstD1* amplification in RT-qPCR, 50 bp
CH507 CH508	TCCTTCATCGTCGACCACAC CTGGCCGAACGACATGTAGT	Tm 60 °C, 1min, 30 cycles	*kstD2* amplification in RT-qPCR, 69 bp
CH509 CH510	GCTACGACCACTACTACGGC CACCACCTTGATCGCGTAGA	Tm 60 °C, 1min, 30 cycles	*kstD3* amplification in RT-qPCR, 92 bp
CH518 CH519	GGCATCTCACTGGGCGAAT CCAGGGAGACGGGACAGTA	Tm 60 °C, 1min, 30 cycles	*D092_14375* reference gen in RT-qPCR, 69 pb
CH371 CH372	GCACATGGCCGGCGCGTAA GGTCGAGGGTGCCCGTTCGG	Tm 58 °C, 0.5 min, 25 cycles	*kstD1 – MFS transporter* coexpression (RT-PCR) 210 bp
CH88 CH337	CTCCGTCCGCCGAACAGAAAG CGGTGGCGAAGGTGGGCAG	Tm 58 °C, 0.5 min, 25 cycles	*kstD3 –hsd4B* coexpression (RT-PCR) 210 bp
CH338 CH339	TCCTCACCCTCGCCTGACCC GGGGTTCGGTGCGAATCGTC	Tm 55 °C, 0.5 min, 25 cycles	*hsd4B*-choG coexpression (RT-PCR) 158bp

Primer	Sequence	PCR conditions	Use
CH259 CH271	**AACATA**TGGTGGATTGGGCAGAGGAAT **GATAT**CGACGACGGCATGGTAGTAGGT	Tm 50 °C, 0.5 min, 25 cycles	*kstD1* amplification in RT-PCR, 247bp
CH53 CH263	GGTCGCCGTAGTAGTGGTCGTA **GGATCC**GCGACGGGCCCGGCGAGAACAT	Tm 50 °C, 0.5 min, 25 cycles	*kstD3* amplification in RT-PCR, 291bp
CH348 CH349	**TCGCGAGCTAGC**AGGGCGCGCCCAGCTGTC **AGTACTACATGT**AAATCGTAATTATTGGGG	Tm 55 °C, 0.5 min, 30 cycles	*T0-MCS-T1* of pSEVA351 amplification for cloning, 415 bp
CH353 CH354	**GGTACCGCATGCTCTAGA**CTGACCCGATGCCACGGCGCC **TTTAAACTGCAGTCGCGA**TTCCTCTGCCCAATCCACCA	Tm 58 °C, 0.5 min, 30 cycles	Promoter region of *kstD1* and its first 21 nucleotides of coding sequence, amplification for cloning, 478 bp
CH355 CH356	**GGTACCGCATGCTCTAGA**CATCGTGTCCTCCGAAGCGG **TTTAAACTGCAGTCGCGA**CGGTACGGGATTGGTCGCCA	Tm 58 °C, 0.5 min, 30 cycles	Promoter region of *kstD2* and its first 21 nucleotides of coding sequence, amplification for cloning,285 bp
CH357 CH358	**GGTACCGCATGCTCTAGA**CCATGTTCTGCAACCTGTTT **TTTAAACTGCAGTCGCGA**GTATTCCTGCTTGATCATGT	Tm 58 °C, 0.5 min, 30 cycles	Promoter region of *kstD3* and its first 21 nucleotides of coding sequence, amplification for cloning, 142 bp
CH488	**GGTACCGCATG**CATCATTCTATAACGTGT	Tm 58 °C, 0.5 min, 30 cycles	In combination with CH402: last 23 pb of the promoter region of *kstD3* and its first 21 nucleotides of coding sequence, 856 bp
CH423 (R1)	**[Phos]** ACTCGGAGGCCTCGACCAGGAC	RT 55 °C	5'-phosforilated primer for reverse transcription TSS *kstD1*
CH424 (R2) CH425 (F3)	CCGATCCCACGACGAGGACGTC ACCGCCGCCCGTGAGGGACTGT	Tm 55 °C, 0.5 min, 30 cycles	Primers for PCR amplification of circularized *kstD1* cDNA, TSS
CH427 (R1)	**[Phos]** CTCCGTCTTCTCCACGATCAGGG	RT 55 °C	5'-phosforilated primer for reverse transcription TSS *kstD2*
CH428 (R2) CH429 (F3)	CCCGACCCGATCACGAGCAGGT TCACCGCGCACGAGCTGGGGTT	Tm 55 °C, 0.5 min, 30 cycles	Primers for PCR amplification of circularized *kstD2* cDNA, TSS
CH443 (R1)	**[Phos]** CGGAATCCACACGCCGCCGCCCGA	RT 58 °C	5'-phosforilated primer for reverse transcription TSS *kstD3*
CH431 (R2) CH432 (F3)	GGTCAGCGCGGCGGTCATTC GCGTCGTCCTGGTCGAGAAGGC	Tm 56 °C, 0.5 min, 30 cycles	Primers for PCR amplification of circularized *kstD3* cDNA, TSS
CH401 CH402	**CCGCGGTCTAGATCGCGA**GTGCAATACGAATGGCGAAAA **GGTACCAGATCTAAGCTT**TCATCAGCCAATCGACTGGCGAGC	Tm 58 °C, 0.5 min, 30 cycles	*Am*^{r} (*Nru*I-*Hind*III) amplification for cloning 804pb

*Nde*I-*Bam*HI (*kstD3*), *Nde*I-*Bgl*II (*kstD2* and *kstD1*) restricted PCR products were first cloned into pGEM-T Easy Vector (Promega) and then moved to the shuttle vector pTip-QC1 (Nakashima and Tamura 2004). Expression plasmids either with or without a *kstD* ORF were used to

electroporate *Rhodococcus erythropolis* CECT3014. Selection was made in LB with 34 µg/mL Cm.

Cultures of *R. erythropolis* harbouring expression plasmids (pTip-QC1, pTip-KsTD1, pTip-KstD2 or pTip-KstD3) were grown (30 °C, 200 rpm) in 50 mL LB broth supplemented with Cm until an OD_{600nm} of 0.6-0.8. After that, expression was induced by adding 1 µg/mL thiostrepton (Sigma) to the culture. Cells were kept growing for 24 h more and they were collected at 5000 rpm for 10 min, washed twice in 50 mM phosphate buffer pH 7.0, concentrated to 5 mL and sonicated. The resulting cell extracts were used for analysis of KstD activity with a range of steroid substrates. Total protein content was measured by Bradford assay (Bradford 1976) and protein expression was followed by collecting 200 µL samples, cell-pelleting at 10000 rpm for 2 min and resuspending in 200 µL of 50 mM phosphate buffer pH 7.0 Samples were analysed by PAGE-SDS in 12.5% w/v gels and using 10 µg of total protein per lane.

2.3. KstD enzymatic assay

The kinetic of the enzymatic extracts was determined as previously stated (van der Geize, Hessels *et al.* 2002, Knol, Bodewits *et al.* 2008). The cell-free extracts were incubated with AD (Sigma), 9OH-AD (Organon Biosciences), 17β-hydroxy-5α-androstan-3-one or 5α-Tes, DOC (deoxycorticosterone), corticosterone, testosterone, progesterone, 19OH-AD (Sigma), 4-BNC (4-pregnen-3-one-20β-carboxylic acid) (Steraloids), 5β-androstane-3,17-dione (Steraloids), 4-cholestene-3-one (Sigma) or 5α-cholestan-3-one (Acros Organics). Enzyme activities were measured spectrophotometrically at 30 °C using 2,6-dichlorophenol-indophenol (DCPIP, Sigma) as an artificial electron acceptor. The reaction mixture (1 mL) consisted of 50 mM Tris (pH 7), 80 µM DCPIP, cell-free extract and 200 µM steroid in ethanol (methanol in the case of 5α-cholestane-3-one). Four replicates were analysed. Activities are expressed as mean values ± SD in units per milligram of protein; one unit is defined as the amount of enzyme which causes the reduction of 1 µmol of DCPIP/min (ε_{600} = 21 $mM^{-1}cm^{-1}$) after taking into account the value of the activity of control (cells harboring the empty pTip-QC1 vector) for a certain steroid.

Total protein concentration (mg/mL) was measured by Bradford assay (Bradford 1976). The kinetics of the KstD enzymes were determined by incubating the cell-free extracts with varying concentrations of steroid substrates. The kinetics parameters were analysed by nonlinear regression curve fitting of the data to the Michaelis- Menten equation using Hyper32 1.0 software (Informer Technologies, http://hyper32.software.informer.com/).

2.4. Expression analysis by RT-PCR and RT-qPCR

RNA samples for RT-PCR experiments were obtained from mid-log exponential phase cultures (OD_{600nm} 0.7-0.8). Total RNA was prepared with the RNeasy Mini Kit (Qiagen) following the manufacturer's indications with the following modification: 50 mg of acid-washed glass beads (150 µm diameter) were added in the first step and each sample was shaken at maximum speed in a Bullet Blender for 5 minutes. The cell debris was removed by centrifugation. The supernatant was subjected to the RNeasy Mini Kit (Qiagen) protocol. The total RNA obtained (0.5-1 µg) was treated once with 5 U of Turbo DNase RNase-Free (Ambion) in a 700 µL volume for 2 hours at 37 °C. RNA samples were extracted with 1 volume of acid phenol (Sigma), vigorously shaken and incubated at room temperature for 15 minutes. After 15 minutes centrifugation, the upper phase was precipitated by addition of 0.12 volumes of 5 M NH_4Ac, 0.02 volumes of glycogen (5 mg/mL) and 1 volume of isopropanol, washed twice in 70% ethanol and dissolved in water. Samples were treated with DNase until no DNA was detected by PCR to avoid DNA contamination. The RNA concentration was evaluated using a NanoDrop Spectrophotometer ND-1000. For the RT-PCR the cDNA was synthesized using SuperScrip III Reverse Transcriptase (Invitrogen) following the manufacturer's indications. cDNA was used as template (25 ng) for PCR reactions (20 µL final volume). Controls without reverse transcriptase (RT-) were used to detect any contamination of undigested DNA in the RNA preparations. PCR products were analysed in 0.8% agarose gels.

To quantify the expression of the three KstD genes, a RT-qPCR analysis was performed using RNA from wild-type strains cultured in M457 minimal medium containing the desired carbon and energy source (2 g/L sodium acetate, 0.44 g/L AD or 0.6 g/L cholesterol). The RNA quality was assessed by using Bioanalyzer 2100 (Agilent). cDNA was synthetized using 1µg of RNA with the High Capacity RNA to cDNA Kit (Applied Biosystems). The RT-qPCR analysis of cDNA was performed on Applied Biosystems QuantStudio 12K Flex Real-Time PCR Systems. The reaction conditions were 10 min at 95 °C followed by 40 cycles of 15 seconds at 95 °C and 1 min at 60 °C for extension. The temperature of the melting curve was from 60 to 95 °C. The gene D092_14375 was used as an internal control to normalize messenger RNA levels. All reactions were performed in triplicate. The RT-qPCR experiment and the analysis of the relative fold difference of each gene using the $2^{-\Delta\Delta Ct}$ algorithm was performed in the Genomic Unit of Universidad Complutense de Madrid. The sodium acetate grown culture was used as the reference medium. Therefore, the relative expression indicates how many times the expression level of a certain gene is detected respect to the levels detected when growing on sodium acetate.

2.5. *In silico* analyses

DNASTAR (Lasergene) programs were used to analyse sequences and to design primers. The *R. ruber* strain Chol-4 genomic DNA has been previously sequenced (Fernández de las Heras, Alonso *et al.* 2013). BioEdit program was used to perform local-blast alignments within the genome data (NCIB: ANGC01000000). Putative signal peptides were predicted by SignalP 4.1 server using a model trained on Gram-positive bacteria (Petersen, Brunak *et al.* 2011). Sigma 70 putative promoters predictions were performed using the BPROM server, a bacterial sigma70 promoter recognition program with about 80% accuracy and specificity (Solovyev and Salamov 2011), the Neural Network Promoter Prediction (NNPP) based on prokaryotes (Reese 2001) in all cases with a score value of ≥80%, or the webserver PePPER for prediction of prokaryote promoter elements and regulons (de Jong, Pietersma *et al.* 2012).

For the protein modelling we employed different software. I-Tasser (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) was used for an approach to protein structure and function prediction (Yang, Yan *et al.* 2015) and PredictProtein (www.predictprotein.org/home) was used for the secondary structure, solvent accessibility and transmembrane helix prediction (Yachdav, Kloppmann *et al.* 2014). COBALT was used as a multiple sequence alignment tool to find similarities among the catalytic residues (www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi) (Papadopoulos and Agarwala 2007) and PyMOL (www.pymol.org/) as a molecular visualization system (PyMOL Molecular Graphics System, Version 0.99, Schrödinger).

2.6. Promoter cloning and characterization

The *Nhe*I-*Pci*I 0.4 Kb multiple cloning site (mcs) from pSEVA351 (Silva-Rocha, Martinez-Garcia *et al.* 2013, Durante-Rodriguez, de Lorenzo *et al.* 2014) was cloned into pNV119 vector (Chiba, Hoshino *et al.* 2007), from now on named as pNVS .

The putative *kstD* promoter sequences were amplified by PCR, from the end of the upstream flanking gene to the end of the first six amino acid codifying sequence. The *Xba*I, *Pst*I-flanked *kstD* promoter regions (*kstD1p, kstD2p* and *kstD3p*) and the *Kpn*I, *Pst*I-flanked *kstD3*^{b}*p* minimal promoter were cloned into pNVS. The resulting vectors were designated pNVSP1, pNVSP2, pNVSP3 and pNVSP3^{b}, respectively.Apramycin resistance gene (Am^{r}) was amplified by PCR from plasmid pIJ773 (Gust, Challis *et al.* 2003), from start to stop codon. The *Nru*I/*Hin*dII digested Am^{r} fragment was cloned in each one of the previous *kstD* constructions. The resulting plasmids were

checked by sequencing (Secugen) and named pNVSP1-A, pNVSP2-A, pNVSP3-A and pNVSP3^{b}-A, respectively. All the primers used are listed in Table 4. PCR was performed under standard conditions using High Fidelity PCR (Roche) with glycerol 5% and a basic program unless stated otherwise.

As a control, a plasmid without any promoter but carrying the Am^{r} gene (pNVSA) was made by digesting pNVSP1-A with *Nru*I-*Xba*I; the resulting 4.2 Kb fragment was cut from a 1% agarose gel and purified with GENECLEAN Turbo Kit. Blunt ends suitable for ligation with T4 DNA ligase (Takara) were generated using the End repair kit (DNA terminator, Lucigen). The final ligated product was used to transform *E. coli* DH5αF'. Deletion was checked by sequencing (Secugen). The plasmid set was introduced in *R. ruber* strain Chol-4 by electroporation (200 or 400 µL cells with 1 µg DNA at 400 Ω, 25 mA, 2.5 µF; 10-11 milliseconds), the resulting cells were suspended in 800 µL of LB and kept for 6 min at 46 °C, and then for 5 hours at 30 °C without shaking. Finally, they were plated on LB Agar with 200 µg/mL kanamycin and kept at 30 ºC. To verify the presence of plasmids, two colonies of each plate were picked and grown in 3 mL of LB-200 µg/mL kanamycin. Plasmid were extracted using the method described in Hopwood *et al.*, 1985 (Hopwood, Bibb *et al.* 1985) and used to transform *E. coli* strain DH5αF' (Sambrook, Fritsch *et al.* 1989). Plasmids obtained from *E. coli* colonies grown at 37 °C in 50 µg/mL kanamycin were verified by sequencing (Secugen). Finally, those colonies of *R. ruber* strain Chol-4 harbouring the right recombinant plasmids were picked and grown in agar minimal media at 30 °C with different carbon source with or without apramycin (300 µg/mL) or kanamycin (200 µg/mL).

To define the transcriptional start sites (TSS) of *kstD* promoters, a transcription start point protocol (ARF-TSS) (Wang, Lee *et al.* 2012) was used on *R. ruber* cells with the following modifications.RNA from *R. ruber* cells growing in different carbons sources cultures was isolated. The culture media were: LB for *kstD1* TSS, minimal medium supplemented with AD for *kstD2* TSS and minimal medium supplemented with cholesterol for *kstD3* TSS. Total RNA was isolated as described previously (Fernández de las Heras, Perera *et al.* 2014). It was qualified by electrophoresis and quantified by Nanodrop 1000 (NanoDrop Technologies).

20 µg of RNA were reverse transcribed with a gene specific phosphorylated 5'-end primer (R1 for each *kstD*) using SuperScrip III Reverse Transcriptase from Invitrogen. As a result, cDNA fragments with the TSS as 3'-end were generated and purified. After removing of RNA with 0.5 µg/µL RNase A (37 °C for 30 min), cDNA was purified using the GENECLEAN turbo Kit (MPI) and final products were treated with T4 RNA ligase from Fermentas (10U, 37 °C for 30 min). This T4 RNA ligase catalyses the ATP-dependent intra- and intermolecular formation of phosphodiester

bonds between 5'-phosphate and 3'-hydroxyl termini of oligonucleotides, single-stranded RNA and DNA. T4 RNA ligase was used in this study to circularize the cDNA. The circularized cDNAs were then used as template for PCR amplification with Expand High Fidelity (Roche) using R2 and F3 primers specific for each *kstD* ORF (see Table 4). PCR products were purified using the GENECLEAN turbo Kit and sequenced (Secugen). The nucleotide upstream the 5'-end of the R1 primer is the transcription initiation site.

3. Materials and methods Chapter 3: Characterization of 3-Ketosteroid 9α-Hydroxylases in *Rhodococcus ruber* strain Chol-4

3.1. Bacterial strains, plasmids and growth conditions

Growth conditions and steroids preparations were made as previously describe in section 1.2 and 2.1. A mix of phytosterols (Gadea S.A.), cholic acid and sodium acetate were directly added to the culture when required. Strains and plasmids used in this study are listed in Table 5.

Table 5: Bacterial strains and plasmids used in this work

Bacteria and plasmids	Description	Reference
Rhodococcus ruber strain Chol-4	Wild type phenotype, Nal^R	CECT7469 (Fernández de las Heras, García Fernández *et al.* 2009)
ΔkshA1, ΔkshA2, ΔkshA3 strain	Single *kshA R. ruber* mutants,Nal^R	This work
ΔkshB strain	Δ*kshB,* Nal^R	This work
ΔkshA1,2 or ΔkshA1,3 or ΔkshA2,3 strain	Double *kshA R. ruber* mutants, Nal^R	This work
ΔkshA1,2,3 strain	Triple *kshA R. ruber* mutant, Nal^R	This work
E. coli DH5α	F' *endA1 hsdR17* (r_K^- m_K^+) *glnV44 thi-1 recA1 gyrA* (Nal^R) *relA1* Δ(*lacIZYA-argF*) *U169deoR* (ϕ80*dlac*Δ(*lacZ*)*M15*)	Laboratory collection
E. coli S17.1	*recA pro hsdR RP4-2-Tc::Mu-Km::Tn7*	(Simon, Priefer *et al.* 1983)
pBluescript II KS^+	Cloning vector, Ap^R	Stratagene
pNV119	Km^R, *Nocardia-E. coli* replicative shuttle vector	(Chiba, Hoshino *et al.* 2007)
pK18mobsacB	Km^R RP4*mob*, mobilizable cloning vector with *sacB* gene for positive selection	(Schafer, Tauch *et al.* 1994)
pK18(*ksh*A1U+D)	pK18mobsacB harbouring a *Eco*RI-*Hind*III genomic fragment containing a *kshA1* truncated ORF	This work
pK18(*ksh*A2U+D)	pK18mobsacB harbouring a *Xba*I-*Hind*III genomic fragment containing a *kshA2* truncated ORF	This work
pK18(*ksh*A3U+D)	pK18mobsacB harbouring a *Hin*dIII-*Xba*I *R. ruber* strain Chol-4 genomic fragment containing a *kshA3* truncated ORF	This work
pK18(*ksh*BU+D)	pK18mobsacB harbouring a *Eco*RI-*Pst*I *R. ruber* strain Chol-4 genomic fragment containing a *kshB* truncated ORF	This work

3.2. DNA preparation, sequencing and *in silico* analysis

Chromosomal DNA extraction from *R. ruber* strain Chol-4 was performed as previously described in section 1.1. The BioEdit program vs 7.1.3.0 (Hall 1999) was used to blast a known ORF against the published genome of *R. ruber* (Fernández de las Heras, Alonso *et al.* 2013). The GenBank accession number for the *R. ruber* genome nucleotide sequences reported in this paper corresponds to NZ_ANGC00000000.1.

The ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/), the RAST server (Aziz, Bartels *et al.* 2008) and the pDraw32 (http://www.acaclone.com/) programs were used to analyse the ORFs contained in the corresponding contigs. Putative domains in the ORFs were searched with the BLASTs server (http://expasy.org/prosite/). DNASTAR (Lasergene) programs were used to analyse sequences, to elaborate cladograms (MegAlign program using the alignment with the Clustal W method) and to design primers. The protein modelling was done as described in section 2.5.

3.3. PCR, RT-PCR analysis and RT- qPCR

For RT-PCR analysis, Chol-4 cultures were grown on 50 mL of 457 DSMZ medium with either 0.6 mg/mL cholesterol (added as a powder) or 2 mg/mL sodium acetate. Cells were harvested in mid-exponential growth phase ($OD_{600\ nm}$ = 0.6) after addition of 5 mL of a 9:1 (v/v) mix of ethanol/phenol and further centrifugation at 5000 rpm, 4 °C for 15 min. RNA extraction and RT-qPCR assay was made as described here in section 2.4.

PCR was performed under standard conditions using Expand High Fidelity PCR System Enzyme Mix (Roche). For amplifications of Chol-4 DNA, glycerol at a final concentration of 5% was also added to the reaction mix. Primers, dimethyl sulfoxide, dNTPs and template DNA were added at a final concentration of 0.4 µM, 4% (v/v), 200 µM and 10 ng/µL respectively. PCR conditions were: 30 cycles of 1 min at 95 °C, 1 min at the desired Tm and 0.5-3 min at 72 °C unless stated otherwise.

Primers used in this work are listed in Table 6.

Table 6: Primers and PCR conditions use in this work

Primer	Sequence	PCR conditions	Use
CH213 CH214	**GAATTC**CCTTCGTCGTCGAGCCCTG **TCTAGA**CCCGTGGCGGCGTGGATGAA	Tm 55 °C, 1.5 min, 30 cycles	*kshA1* up-flanking region *Eco*RI-*Xba*I amplification
CH215 CH216	**TCTAGA**GCCGAAGGCGTGGACCGAATGCG **AAGCTT**CGGCGCACGAGAGGGAGGATA	Tm 55 °C, 1.5 min, 30 cycles	*kshA1* down-flanking region *Xba*I-*Hin*dIII amplification
CH209 CH227	**TCTAGA**CCCGGGCGCTCGTCCAT AG**GATATC**GCGCGGTGGGAGCAGCAGA	Tm 58 °C, 1.5 min, 30 cycles	*kshA2* up-flanking region *Xba*I-*Eco*RV amplification
CH225 CH226	GG**AAGCTT**CCCGCCGCCGAAGATGTAG G**GATATC**GCGGACTTCATCGTGGGGTGTG	Tm 58 °C 58 °C, 1.5 min, 30 cycles	*kshA2* down-flanking region *Eco*RV -*Hind*III amplification
CH228 CH229	CC**AAGCTT**GCCGGCGCCGAACATCACG **CCAGATC**TCGTCGAGGATCTTCAGGTC	Tm 58 °C, 1.5 min, 30 cycles	*kshA3* up-flanking region *Hin*dIII-*Bgl*II amplification
CH230 CH231	**TCAACT**GCCACTACCCCGTCACAC **GTCTAGA**CATCGCTGTGACGGCATCTGACTTC	Tm 58 °C, 1.5 min, 30 cycles	*kshA3* down-flanking region *Bgl*II-*Xba*I amplification
CH201B CH202	**GAATTC**GCAGGTGAGAGGCGGTTTTTAC **AAGCTT**GACTCGACGTGCCCTACTCCT	Tm 55 °C, 1.5 min, 30 cycles	*kshB* up-flanking region *Eco*RI-*Hin*dIII amplification
CH203 CH204	**AAGCTT**GTACGCGAACTTGTCCTGGAG **CTGCAG**GGTGTTCATCCGGTTCTTCTCGTA	Tm 55 °C, 1.5 min, 30 cycles	*kshB* down-flanking region *Hin*dIII-*Pst*I amplification
CH367 CH368	TCCACTGGACGGAGGTGTGC CCCGTGGCGGCGTGGATGAA	Tm 58 °C, 0.5 min, 25 cycles	*kshA1*- HP (Hypothetical protein) co-expression 159 bp
CH369 CH370	GCCGATCATCCAGTAAGTCA ACCTTCTCGGCGGTCTCCTC	Tm 58 °C, 0.5 min, 25 cycles	*HP – dehydrogenase* co-expression (RT-PCR) 193 bp
CH227 CH266	**AGGATAT**CGCGCGGTGGGAGCAGCAGA **AAGCTT**TCGAGCCGGCCTCCTTCATCTT	Tm 58 °C, 0.5 min, 25 cycles	*kshA2 –hydroxylase* co-expression (RT-PCR) 241 bp
CH328 CH364	TCGTCCATGCGCAGGTACA TGCGGGCGATGATCTGACC	Tm 55 °C, 0.5 min, 25 cycles	*hydroxylase-dioxygenase* co-expression (RT-PCR) 184 bp
CH365 CH366	GTCGGCACGGGTGCTCCTC TCGGGGTGATGTCCTGATCG	Tm 55 °C, 0.5 min, 25 cycles	*dioxygenase – lclR* co-expression 256 bp
CH206 CH373	ACTAGTTCGCCGACGTCACCGAGAAGA TGATCTTGTCGATGGCGGTC	Tm 55 °C, 0.5 min, 25 cycles	*kshA3 –HP* co-expression 241 bp
CH374 CH375	ATCACGATTCCGGACTGACA TGAACTGGTTGAATTCGCGG	Tm 55 °C, 0.5 min, 25 cycles	*HP-HP* coexpression 241 bp
CH469 CH470	TACGAGGGCTACGACATC TACATCAGGACGAACTTGTT	Tm 60 °C, 1min, 30 cycles	*kshA1* amplification in RT-qPCR, 52 bp
CH522 CH523	AAGACGTTCGAGCAGGACAT GTTGTCGATCTTGGCCTTGT	Tm 60 °C, 1min, 30 cycles	*kshA2* amplification in RT-qPCR, 50 bp
CH524 CH525	ACGCTGGTACGAGCAGTTCT GTCGACCTCGTACTCGAAGC	Tm 60 °C, 1min, 30 cycles	*kshA3* amplification in RT-qPCR, 79 bp
CH518 CH519	GGCATCTCACTGGGCGAAT CCAGGGAGACGGGACAGTA	Tm 60 °C, 1min, 30 cycles	*D092_14375* reference gen in RT-qPCR, 69 pb

3.4. Mutagenesis of the *kshA* and *kshB* genes of *R. ruber* strain Chol-4.

Primers and conditions used in this experiment are summarized in Table 6. Unmarked gene deletions were carried out as described previously in *R. erythropolis* SQ1 involving conjugative transfer of a mutagenic plasmid carrying the *sacB* selection system (van der Geize, Hessels *et al.* 2001).

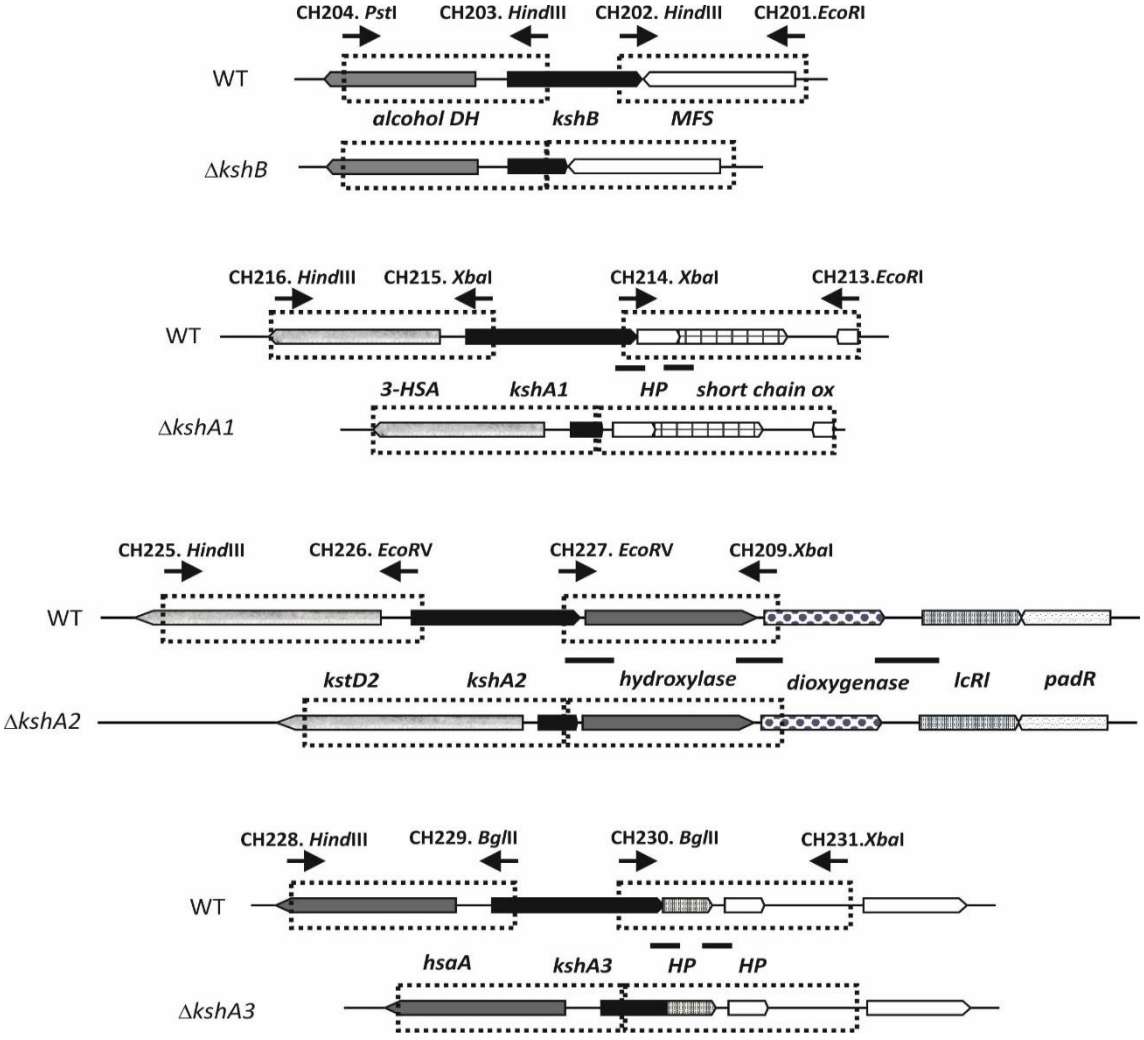


Figure 1. Scheme of the *kshA* and *kshB* ORFs deletion in *R. ruber* strain Chol-4 genome. Solid lines indicate co-transcription of the ORFs confirmed by RT-PCR. *HP*: hypothetical protein; *DH*: dehydrogenase; *ox*: oxidase; *MFS*: major facilitator superfamily transporter; *padR* and *lclR*: transcriptional regulators.

Specific pairs of primers were designed from up and downstream sequences of the terminal regions of each ORF in order to get PCR amplicons from isolated *R. ruber* strain Chol-4 genomic DNA. The primers also contained suitable restriction sites to allow the later cloning and fusion of both terminal regions of every ORF. The amplified *kshA1* terminal regions contained *Eco*RI-*Xba*I and *Xba*I-*Hin*dIII restriction sites in the up and down terminal regions respectively; *Xba*I-*Eco*RV (up) and *Eco*RV-*Hin*dIII (down) were in the amplified *kshA2* terminal regions; *Hin*dIII-*Bgl*II

(up) - *Bgl*II-*Xba*I (down) were in the amplified *kshA3* terminal regions and *Eco*RI-*Hin*dIII (up) - *Hin*dIII-*Pst*I (down) in the amplified *kshB* terminal regions. These amplicons were cloned into pGEM-T-Easy vectors and every pair was combined in order to get a DNA fragment containing a truncated *kshA* or *kshB* ORF that keeps the right reading frame to avoid polar effects (Figura 1). These DNA fragments were cloned into the pK18mobsacB plasmid (Schafer, Tauch *et al.* 1994) to construct the mutagenic plasmids pK18(*ksh*A1U+D), pK18(*ksh*A2U+D), pK18(*ksh*A3U+D) and pK18(*ksh*BU+D), which were used to delete *R. ruber* strain Chol-4 *kshA1*, *kshA2*, *kshA3* and *kshB* genes, respectively.

The mutagenic plasmids were introduced into *E. coli* S17.1 and mobilized to *R. ruber* strain Chol-4 by conjugation as already published (Fernández de las Heras, van der Geize *et al.* 2012).

kshA and *kshB* ORF truncation were then achieved as a result of a spontaneous double homologous recombination process within the genome of *R. ruber* strain Chol-4. Deletions were checked by colony PCRs. Double or triple mutations were generated taking the single or double mutant strain as a new recipient, respectively.

3.5. Monitoring of steroids intermediaries by thin layer chromatography (TLC) in *ΔkshA* and *ΔkshB* on growing or resting-cells.

The protocol for resting cells was as followed. Chol-4 wild-type and *ΔkshA* and *ΔkshB* strains were grown in 100mL of LB medium in 500 mL flasks at 30 °C in a rotary shaker (250 rpm) for 2 days. One hour before harvesting, chloramphenicol was added to a final 34 µg/mL. The equivalent of cells to 30 mL of culture at OD_{600nm} of 18, were harvested at 4˚C by centrifugation at 10,000 x g for 15 min, washed twice with 0.85% NaCl, and suspended in 25 mM of HEPES (Sigma) pH 7. Cholesterol (1.7 mg/ml) and AD (1 mg/ml) were added as powder. The resting-cell reaction was allowed to proceed at 30°C and 150 rpm. Aliquots of 500 µL were taken at fixed times and their lipid fraction was obtained by extraction twice with 1 mL chloroform and let drying at 65°C.

50 µL chloroform were added to every dried sample; an aliquot of 10-20 µL of the solution (in the case of Cholesterol) and 10 µl (in the case of AD) were applied on aluminium TLC Silica gel 60 F254 sheets (Merck). Chromatography was performed using a mix of hexane:ethylacetate (10:4, v/v) as solvent and spots were revealed by UV exposure (λ_{254nm}). Afterwards, the TLC plate was dunked in a sulfuric acid:methanol solution (1:9, v/v), dried with warm air and heated 2 min at 100 °C. 1 µg of cholesterol, AD, ADD and 9OH-AD standard control samples were also included.

To monitor intermediary production in culture, mutants were grown in 3 mL LB for two days. Then cells were washed 2–3 times with minimal medium, 50 µL of the washed cells suspended in 50 mL of minimal medium supplemented with cholesterol (1.7 mg/ml) or AD (1 mg/ml) and kept growing at 30 °C, 250 rpm. Aliquots of 2 mL were taken from the culture at fixed times after $OD_{600\ nm}$ was measured. Their lipid fraction was obtained by extraction and TLC analysis was done as previously described.

4. Materials and methods Chapter 4: Biotechnological applications of the genus *Rhodococcus*: generation of a dehydrogenation strain

4.1. Bacterial strains, plasmids and growth conditions

Growth conditions and steroids preparations were made as previously describe in section 1.2 and 2.1. Strains and plasmids used in this study are listed in Table 7.

Table 7: Bacterial strains and plasmids used in this work

Bacteria and plasmids	Description	Reference
Rodococcus erythropolis strain *3014*	Wild type, Nal^r	CECT3014
E. coli DH5α	F' *endA1 hsdR17* ($r_K^- m_K^+$) *glnV44 thi-1 recA1 gyrA*(Nal^r) *relA1* Δ(*lacIZYA-argF*) *U169deoR* (ϕ80*dlac*Δ(*lacZ*)*M15*)	Laboratory collection
pMV261MSGM_5941	Expression and shutlle vector *Mycobacteria*l –*Escherichia coli* with *MSGM_5941* gene, Km^r	(Galán, Uhía *et al.* 2016)
pGEMT-EasyStdH	Cloning vector *E. coli*, with *stdH* gene of *Pseudomonas sp*. , Ap^r.	Laboratory collection
pTip-QC1	Shuttle vector *E. coli-Rhodococcus*, Ap^r, P_{TIPA} Chl^rREPAB (pRE2895)	(Nakashima and Tamura 2004)
pTIP-5941	pTP-QC1 vector with *MSGM_5941 gene*	This work
pTIP-StdH	r pTP-QC1 vector with *stdH gene*	This work

Primers employed for cloning and verification used in this work are listed in Table 8.

Table 8: Primers and PCR conditions use in this work.

Pirmer	Sequence	PCR conditions	Use
CH495 CH496	TTATAA**CATATG**ACTGGACAGGAGTAC TTGATC**AAGCTT**TCAGCCCTTCCGGGAGGC	Tm 55 °C, 1 min, 30 cycles	PCR for cloning *MSMEG_5941, restriction sites NdeI-HindIII*, 1.8kb
pTIP_Fw pTIP_Rv	CGGGCTTGCACCTCACGTCA GTGTTTGTGCAGGTTTCGCG	Tm 58 °C, 1 min, 30 cycles	Sequencing primers for pTIP-QC1 cloning

The restriction sites are indicated in bold.

4.2. Cloning and heterologous expression of *MSMEG_5941* in *R. erythropolis* 3014

M. smegmatis gene *MSMEG_5941* was amplified by PCR performed under standard conditions. The enzyme High Fidelity PCR (Roche) was used at 30 cycles of 1 min at 95 °C, 1 min at Tm 55 °C and 2 min at 72 °C. PCR products were digested with the *Nde*I-*Hind*III restriction enzymes. The *stdH* gene of *Pseudomonas putida* was obtained from digestion of the plasmid pGEMT-EasyStdH, with *Nde*I-*Hind*III restriction enzymes. All genes were cloned into the pTip-QC1 expression plasmid (Nakashima and Tamura 2004). Plasmids obtained, plus a plasmid without insert (pTIP-QC1, pTIP-StdH and pTIP-5941) were used to electroporate *R. erythropolis* CECT3014 according to the protocol previously described in section 2.6. The colonies selection were made in LB with Cm 34 µg/mL. *R. erythropolis* harboring pTIP-QC1 and pTIP-5941 plasmids were grown in 50 mL of LB supplemented with Cm 34 µg/mL until OD600nm of 0.6-0.8. Then the expression was induced by adding 1 µg/mL thiostrepton (stock 50 mg/mL) to the culture. Cells were kept under growth for a further 24 hours and harvested at 5000 rpm for 10 min, washed twice in 50 mM phosphate buffer pH 7.0, concentrated in 1 ml and sonicated twice. Expression analysis was performed by PAGE-SDS gel analysis (12.5%, w/v); using 10 µg of total protein per extract and staining was performed with Coomassie Blue.

4.3. $KstD1_{MSMEG}$ enzymatic assay

The enzyme activity was determined as described in section 2.3. The cell free extracts were incubated with the following substrates: testosterone; AD (4-androsten-3,17-dione); 9OH-AD (9α-hydroxy-4-androsten-3,17-dione); Progest (progesterone); 5α-test (5α-testosterone), 19OH-AD (19-hidroxi-4-androsten-3,17-diona); Andros (androesterone); Cort (corticosterone), DOC (deoxicorticosterone); DHEA (dehiroepiandrosterone) y 4-BNC (4-pregnen-3-one-20β-carboxylic acid).

VI. RESULTADOS

CAPÍTULO 1

New insights into *Rhodococcus ruber* Chol-4 genome

1. Background

Rhodococci are known by displaying a big metabolic versatility and by their ability to transform a wide range of pollutant compounds, such as aliphatic and aromatic hydrocarbons, oxygenated and halogenated compounds, nitroaromatics, heterocyclic compounds, nitriles, and various pesticides (Kuyukina and Ivshina 2010).

The analysis of their genomes has revealed a multiplicity of genes, a high genetic redundancy of metabolic pathways, and a complex regulatory network (Alvarez 2010). Moreover, some *Rhodococcus* strains harbor circular and linear plasmids that contain many additional enzymes (Kulakova, Stafford *et al.* 1995; Shimizu, Kobayashi *et al.* 2001; Francis, De Keyser *et al.* 2012). Even the mycolic-acid-containing outer membrane of rhodococci could contribute to the catabolism of many compounds (Iwabuchi, Sunairi *et al.* 2002).

This versatile metabolic capacity and also their environmental persistence and tolerance to stress conditions make *Rhodococcus* strains good candidates for many potential applications in biotechnological processes such as bioremediation, biotransformations or biocatalysis (Martínková, Uhnakova *et al.* 2009; Kuyukina and Ivshina 2010; Yam and Okamoto 2011). Moreover, *Rhodococcus* strains are able to synthesize compounds of industrial interest such as biosurfactants (Kuyukina and Ivshina 2010) and steroid precursors (García, Uhía *et al.* 2012). For all these reasons, the study of the different *Rhodococcus* is necessary to make full use of their biotechnological potential.

Rhodococcus ruber strain Chol-4 that was isolated from a sewage sludge sample, is classified as a Gram-positive bacteria belonging to the actinobacteria taxon with a high guanine-cytosine content (Fernández de las Heras, García Fernández *et al.* 2009). This strain is able to grow in minimal medium supplemented with several compounds, showing a broad catabolic capacity. The draft genome sequence of this bacterium has recently been published (Fernández de las Heras, Alonso *et al.* 2013).

Genome mining performed in different bacterial species has led to correlate the occurrence of certain genes with particular metabolic capabilities. In this work we presented novel results that support the metabolic potential of *R. ruber* strain Chol-4.

2. Results

2.1. General genome features

R. ruber strain Chol-4 genome was sequenced using two different Next Generation Sequencing technologies. First, a genomic library was generated and sequenced in a Roche 454 GS FLX instrument (Fernández de las Heras, Alonso *et al.* 2013). After quality filtering and adapter clipping, this library rendered 242,042 reads with an average length of 400 bp. A second library was generated and processed by pair-end sequencing in an Illumina HiSeq 2000 instrument. This library generated 2,782,965 pair end reads of 90 bp with an average fragment size of 500 bp between pairs. Single-reads from the 454 library and pair-end reads from the Illumina library were combined with four larger sequences that were previously obtained by standard cloning. All the sequences were assembled using SPAdes *de novo* assembler (Bankevich, Nurk *et al.* 2012). The assembly was post-processed using MismatchCorrector, which uses BWA aligner (Li and Durbin 2010) to detect and correct errors. The final assembly contains 50 scaffolds covering 5.4 Mb with a N50 of 438,623 bp and an average read depth above 100X, and it displays a high G+C content of 70.7%.

A comparative genomic analysis performed between this sequence and other genome sequences of *Rhodococcus* described elsewhere is presented in this study. Figure 1a shows the identities between *R. ruber* genome and *R. pyridinivorans* SB3094 (GenBank: NC_023150.1) that could provide an idea of how the *R. ruber* contigs could be arranged.

The genome size of different published *Rhodococcus* is in a range of 3.9 to 10 Mb. These differences in size could be related to the presence of large plasmids and instabilities in the *Rhodococcus* genomes (Larkin, Kulakov *et al.* 2010). The genome of *R. ruber* contains 5.4 Mb, a size that it is within the *Rhodococcus* average size (5.69 Mb: https://www.ncbi.nlm.nih.gov/genome/13525) and that it is quite similar to the *R. equi* ATCC 33707 (5.2 MB), *R. pyridinivorans* SB3094 (5.6 Mb), *R. aetherivorans* or *R. fascians* A44A (both with 5.9 Mb) genome size.

It is important to highlight that we were unable to find any trace of circular plasmids by pulsed-field gel electrophoresis in the genome of *R. ruber* (data not shown).

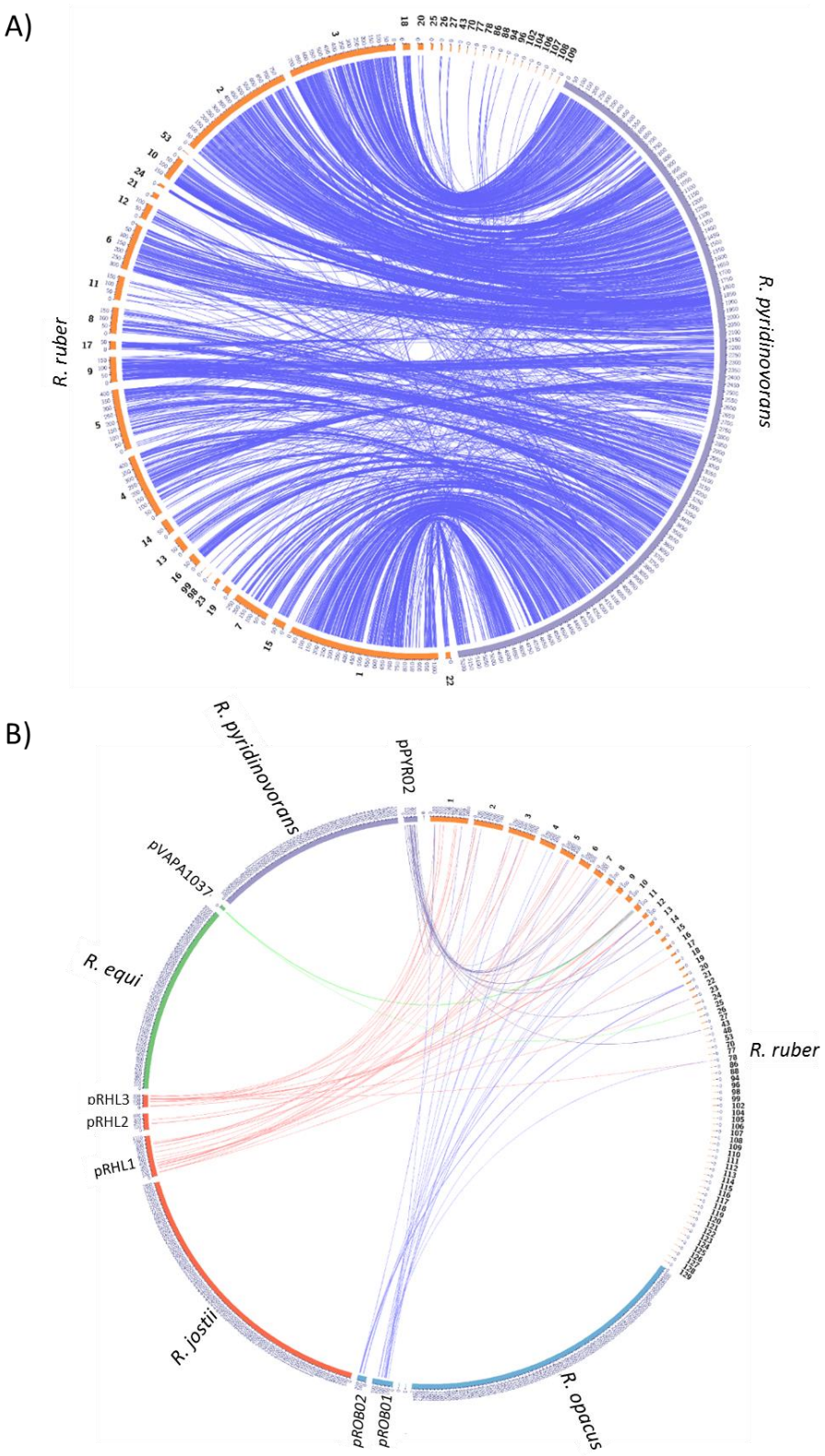


Figure 1. Comparison of *R. ruber* strain Chol-4 genome using CIRCO. Circos plot shows data of *R. ruber* strain Chol-4 genome against: **A)** *Rhodoccus pyridinivorans* genome, **B)** some *Rhodococcus* plasmids.

2.2. Annotation of *Rhodococcus ruber* strain Chol-4

Genome annotation using RAST server (Aziz, Bartels *et al.* 2008) identified 5049 coding sequences and 60 RNAs. Over 53 tRNAS representing 43 different anticodons are encoded in the *R. ruber* genome (Supplemental Table S1). There are at least 7 tRNAS in multicopy: $tRNA^{Met}$ (ATG) is present in 4 copies, $tRNA^{Gly}$ (GGC) in 3 copies and $tRNA^{Val}$ (GTC), $tRNA^{Glu}$ (GAG), $tRNA^{Ala}$ (GCC), $tRNA^{Asp}$ (GAC) and $tRNA^{Leu}$ (CTC) in 2 copies. The codon usage correlates with the high G+C content of this strain as G+C codons are predominant in this organism. Codons that have a T at the third position lack a cognate tRNA in *R. ruber* with the single exception of Arg (CGT). The number of tRNAs is similar to others *Rhodococcus* strains that display a median value of 53 tRNAs although there are exceptions such as *R. rhodnii* ASM72037v1 that contains 69 tRNAs (EMBL: NZ_JOAA00000000.1). There is also a single *rrn* operon located in the contig NZ_ANGC02000002.1 containing the genes for 16S, 23S and 5S rRNA.

The protein-coding sequence detected (up to 5049 ORFs) covers nearly a 91 % of the genome. The analysis of the data suggests that *R. ruber* Chol-4 contains several putative metabolic clusters of biotechnological interest, such as those ones regarding to resistance to several metals or aromatic compounds and steroid catabolism.

The basal metabolism (glycolytic pathways, gluconeogenesis, phosphoenolpyruvate (PEP)–pyruvate-oxaloacetate, the tricarboxylic acid cycle (TCA) and the glyoxylate shunt) based on *R. jostii* RHA1 genome has been proposed (Álvarez 2010). Strain RHA1 seems to be able to use the Embden–Meyerhof–Parnas and Entner-Doudoroff pathways for catabolism of carbohydrates. It is remarkable that *Rhodococcus ruber* Chol-4 contains similar activities within its genome and eventually could share the same pathways.

Among the coding sequences, the analysis revealed that at least: 129 genes are related to the metabolism of aromatic compounds; 15 of them are involved in peripheral degradation pathways (quinate, benzoate and *p*-hydroxibenzoate degradation); 15 genes are related to the aromatic amine catabolism; 6 genes are associated with the gentisate degradation pathway and 4 genes code for the 4 subunits of the anaerobic acetophenone carboxylase 1. It also revealed a small number of genes that are involved in the resistance to antibiotics (resistance to vancomycin, fluoroquinolones, β-lactamase) as well as a relatively large number of genes related to the resistance to toxic compounds, such as mercury and arsenic.

2.3. Mobile elements

There are a few mobile elements within the genome of the strain Chol-4 (Supplemental Table S2), some of them remaining as pseudogenes. Surprisingly, 25% of these mobile elements are concentrated in a 120.000 pb region of contig NZ_ANGC020000011.1.

R. ruber Chol-4 genome has two copies of IS*1164* from the IS*256* transposase family (the prototype of a major family of bacterial insertion sequence elements) in contig NZ_ANGC02000007.1. This element has also been found in other *Rhodococcus* strains (Alvarez 2010). In the same contig, there are two genes (D092_RS18300 and D092_RS17945) that keep a 90-92% bases identity with elements of *Rhodococcus pyridinivorans* SB3094 plasmid (CP006997.1) or with elements of the plasmid pNSL1 (KJ605395.1) from *Rhodococcus* sp. NS1 and a 79% base identity with elements of pMKMS01 (CP000519.1) from *Mycobacterium* sp. KMS.

Other IS elements described for some *Rhodococococcus* were also absent in this draft genome: the IS*2112* element belonging to the IS*110* family that it is found in *R. rhodochrous* NCIMB 13064 and that it is related to genome rearrangements (Kulakov, Poelarends *et al.* 1999) or the IS*1166* elements from the IS*256* family that it is found in *R. erythropolis* IGTS8 (Denome and Young 1995).

Rhodococcus is known to contain systems that favor high-frequency illegitimate recombination processes that could promote the integration of foreign DNA and therefore changes in their genomes (de Vries and Wackernagel 2002). Apart from those derived of mobile elements of unknown functions, strain Chol-4 contains many different recombinases with different roles (Supplemental Table S3).

2.4. Other elements

Some actinobacteria such as *Mycobacterium tuberculosis* contains from 1 to 3 CRISPR elements (CRISPR database, http://crispr.u-psud.fr/)(Botelho, Canto *et al.* 2015). However, *R. ruber* apparently is a strain devoid of detectable CRISPRs systems, similarly to other *Rhodococcus* strains.

On the other hand, it contains a cluster related to specialized protein degradation systems that includes a 20S proteasome activity (subunits α and β), an ATPase (that use ATP to unfold proteins and translocate them into the proteasome) and a system of tagging proteins for degradation

with Pup prokaryotic ubiquitin-like protein; functionally analogous to ubiquitin. Conjugation with Pup serves as a signal for degradation by the mycobacterial proteasome (Figure 2) (Voges, Zwickl *et al.* 1999; Bode and Darwin 2014). Most of the restriction modification systems detected in this genome are classified as type I or IV (Supplemental Table S4).

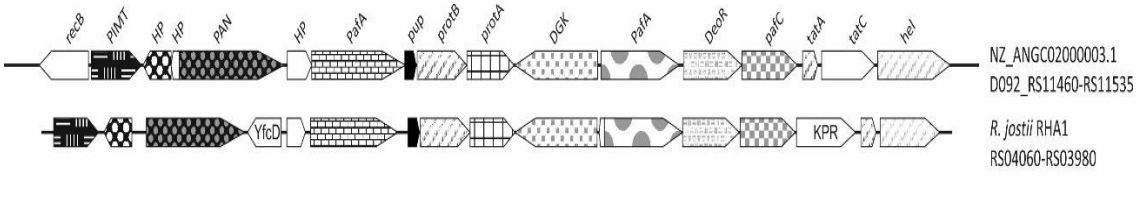


Figure 2. Pup proteasome in *R. ruber*. Abbreviations: **recB**: RecB family exonuclease; **PIMT**: protein-L-isoaspartate methyltransferase (EC 2.1.1.77); **HP**: hypothetical protein; **PAN**: bacterial proteasome-activating AAA-ATPase; **pafA**: proteasome accessory factor, Pup ligase PafA' paralog, possible component of postulated heterodimer PafA-PafA'; **pup**: prokaryotic ubiquitin-like protein Pup; **prota** and **protb**: proteasome subunit α and β (EC 3.4.25.1) bacterial; pup: Prokaryotic ubiquitin-like protein Pup; **DGK**: diacylglycerol kinase; **DeoR**: putative DeoR-family transcriptional regulator; **pafC**: DNA-binding protein; **tatA** and at **tatC**: twin-arginine translocation proteins; **hel**: FIG005666: DEAD/DEAH box helicase; **KPR**: 2-dehydropantoate 2-reductase(EC:1.1.1.169); **YfcD**: nudix hydrolase YfcD (EC 3.6.-.-). All *R. jostii* RHA1 genes have the prefix "RHA1_" that it is not included in the figure.

2.5. Aromatic specific clusters

Actinobacteria are described to be important environmental biodegraders in soil due to their ability to transform a wide range of compounds (Yam, van der Geize *et al.* 2010; Amoroso, Benimeli *et al.* 2013). Aromatic compounds are widely distributed in the biosphere and therefore, the exposure of *Rhodococci* to them makes these strains good aromatic degraders. The catabolism of aromatic compounds is organized in a modular way in *Rhodococcus* strains (Figure 3): the peripheral, the central and the basic pathway. In the peripheral pathway, aromatic compounds (*e.g.* biphenyl and phthalate), are converted into intermediates (*e.g.* catechol and phenylacetate), they are used in central aromatic pathways to obtain common intermediates (*e.g.* tricarboxylic acid cycle metabolites) for the basic pathway (Figure 3) (Yam, van der Geize *et al.* 2010). This kind of organization has been previously named catabolon in other organisms, such as *Pseudomonas* (Luengo, Garcia *et al.* 2001). Consistent with their aerobic lifestyle, the aromatic *Rhodococcus* catabolic pathway is rich in oxygenases (Yam, van der Geize *et al.* 2010).

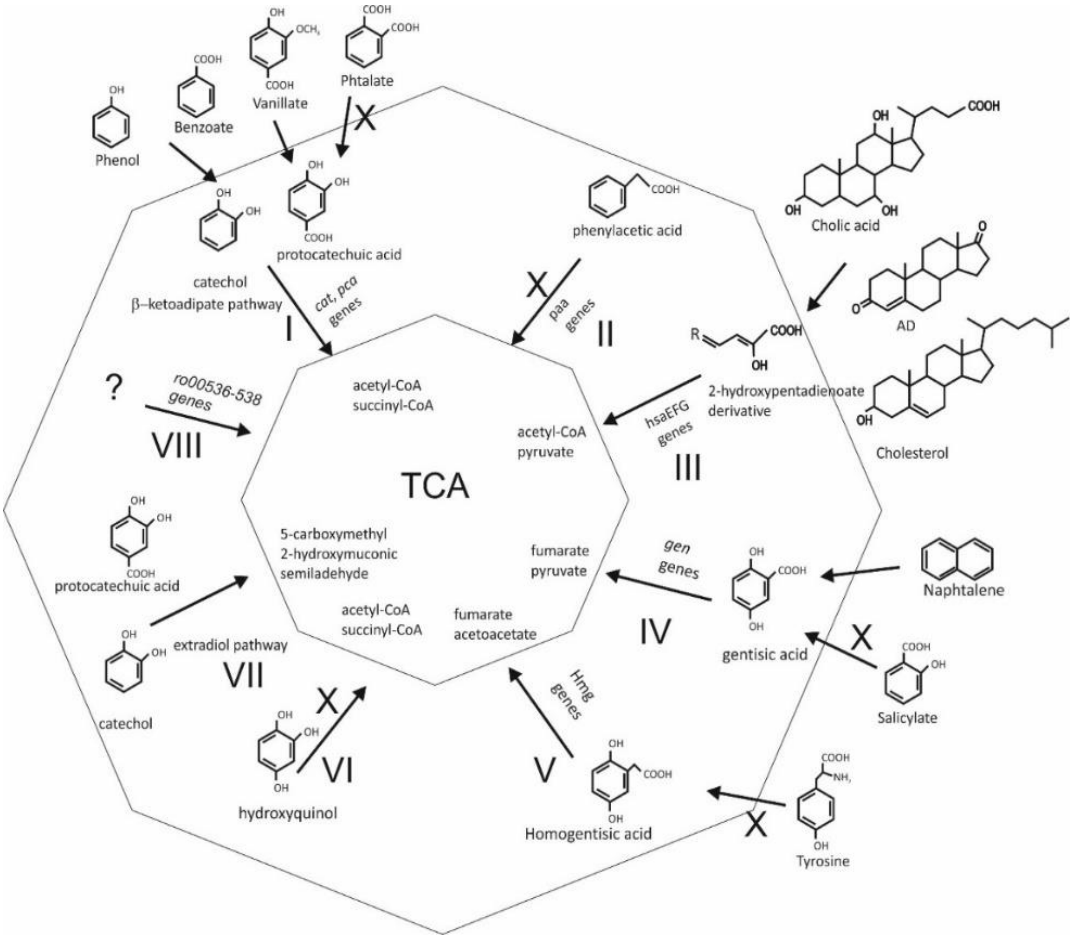


Figure 3. Aromatic central metabolism. Scheme of the aromatic compounds pathways: **I)** β-ketoadipate pathway, **II)** Paa pathway, **III)** 2-hydroxypentadienoate pathway, **IV)** gentisate pathway, **V)** homogentisate pathway (Hmg), **VI)** hydroxyquinol pathwaythe, **VII)** homoprotocatechuate (3,4-dihydroxyphenylacetate) pathway, and **VIII)** a pathway comprising a hydroxylase, an extradiol dioxygenase, and hydrolase found in *R. jostii* RHA1. The symbol "X" indicates the inability of *R. ruber* to grow in the presence of these compounds as single source of carbon and energy.

There are at least eight central aromatic pathways chromosomally encoded in *R. jostii* RHA1 and *R. opacus* B4 (Figure 3 and 4):

I. **The β-ketoadipate**. This pathway is encoded by the *pca* and *cat* genes and is responsible for the conversion of catechol and protocatechuate (3,4-dihydroxybenzoate) to acetyl-CoA and succinyl-CoA by intradiol cleavage of the catecholic intermediate. It is also known as ortho-cleavage pathway (Figure 4 I).

II. **The Paa pathway**. It intervenes in the catabolism of a variety of compounds, including homophthalete, tropate and phenylalkanoates. It involves the derivatization of aromatic acids by CoA, ring hydroxylation by an oxygenase, and a non-oxygenolytic ring fission. The Paa pathway is encoded by the *paa* genes (Navarro-Llorens, 2005) as it is depicted in Figure 4 II.

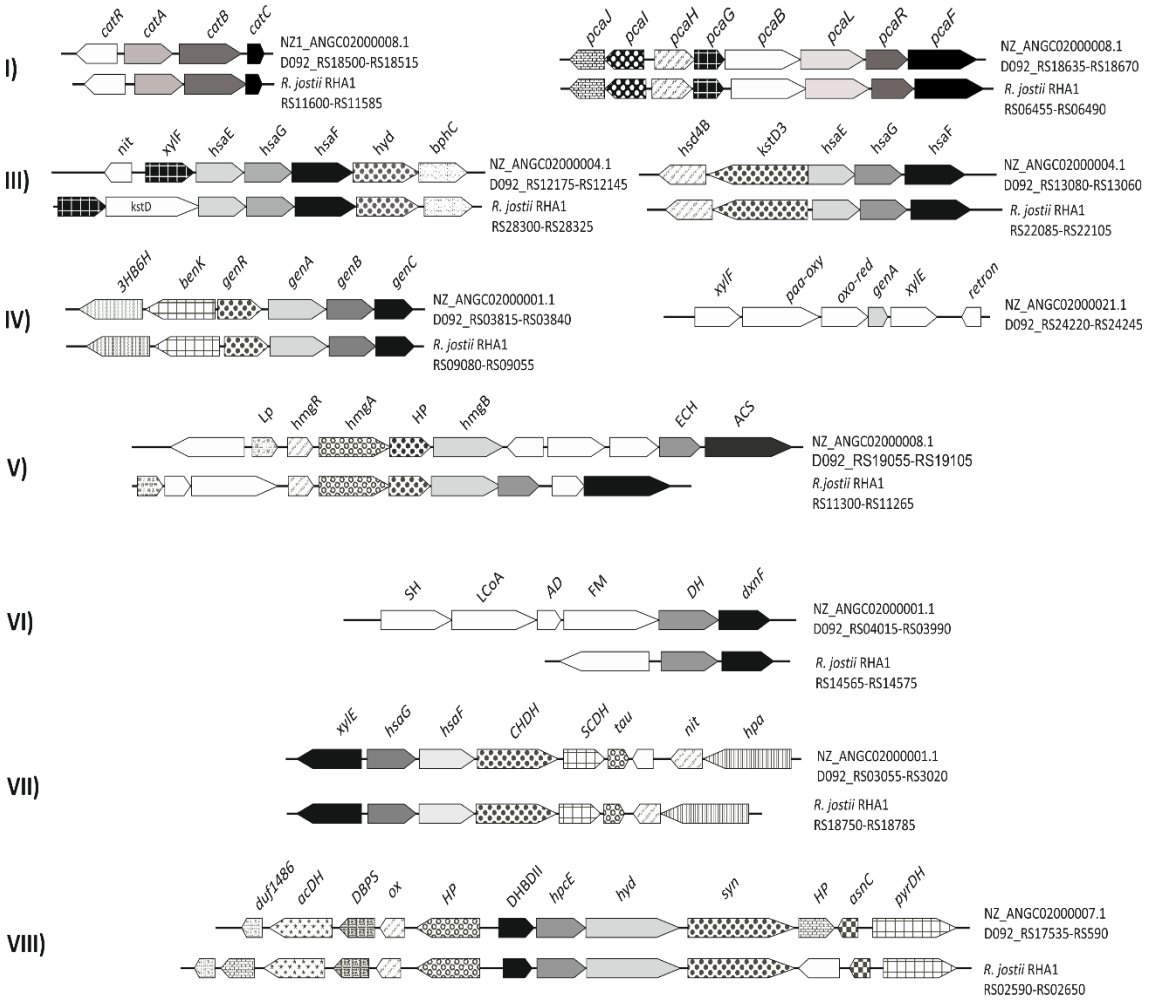


Figure 4. Aromatic clusters detected in *R. ruber* Chol-4 and its comparison with *R. jostii* RHA1.
Abbreviations: **I) ketoadipate pathway:** ***catR***: transcriptional regulator CatR***;*** ***catA***: catechol 1,2 dioyxigenase; ***catB***: muconate cycloisomerase; ***catC***: mucolactone isomerase; ***pcaJ***: succinyl-CoA:3-ketoacid-coenzyme A transferase subunit B (EC 2.8.3.5); ***pcaI***: succinyl-CoA:3-ketoacid-coenzyme A transferase subunit A (EC 2.8.3.5); ***pcaH***: protocatechuate 3,4-dioxygenase β chain (EC 1.13.11.3); ***pcaG***: protocatechuate 3,4-dioxygenase α chain (EC 1.13.11.3); ***pcaB***: 3-carboxy-cis,cis-muconate cycloisomerase (EC 5.5.1.2); ***pcaL***: 4-carboxymuconolactone decarboxylase (EC 4.1.1.44); ***pcaR***: Pca regulon regulatory protein; ***pcaF***: β-ketoadipyl-CoA thiolase. **III) 2-hydroxypentandienoate pathway:** ***nit***: nitrilotriacetate monooxygenase component B (EC 1.14.13); ***xylF***: 2-hydroxymuconic semialdehyde hydrolase (EC 3.7.1.9); ***hsaE***: 2-hydroxypentadienoate hydratase; ***hsaG***: acetaldehyde dehydrogenase, acetylating, (EC 1.2.1.10) it is found in gene cluster for degradation of phenols, cresols, catechol; ***hsaF***: 4-hydroxy-2-oxovalerate aldolase (EC 4.1.3.39); ***hyd***: hydroxylase; ***bphC***: 2,3-dihydroxybiphenyl 1,2-dioxygenase; ***hsd4B***: enoyl-CoA hydratase; ***kstD***: 3-ketosteroid-Δ^1-dehydrogenase (EC 1.3.99.4). **IV) gentisate pathway**: **3HB6H**:3-hydroxybenzoate 6-hydroxylase (EC 1.14.13.1); ***benK*****:** benzoate MFS transporter; ***genR***: transcriptional regulator (IclR family); ***genA***: gentisate 1,2-dioxygenase (EC 1.13.11.4); ***genB***: fumarylpyruvate hydrolase; ***genC***: maleylpyruvate isomerase, mycothiol-dependent (EC 5.2.1.4); ***xylF***: 2-hydroxymuconic semialdehyde hydrolase (EC 3.7.1.9); **paa-oxy**: 4-hydroxyphenylacetate 3-monooxygenase (EC 1.14.13.3); ***oxo-red***:3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100); ***xylE***: catechol 2,3-dioxygenase (EC 1.13.11.2); ***Retron*****:** retron-type RNA-directed DNA polymerase (EC 2.7.7.49). **V) homogentisate pathway**: ***Lp***: uncharacterized protein Rv2599/MT2674 precursor; lipoprotein; ***hmgR***: transcriptional regulator (MarR family); ***hmgA***: homogentisate 1,2-dioxygenase (EC 1.13.11.5); ***HP***: hypothetical protein; ***hmgB***: fumarylacetoacetate hydrolase; ***ECH***: enoyl-CoA hydratase (EC 4.2.1.17); ***ACS****:* acetoacetyl-CoA synthetase (EC 6.2.1.16)/long-chain-fatty-acid--CoA ligase (EC 6.2.1.3). **VI) hydroxyquinol pathway**: ***SH***: salicylate hydroxylase; ***LCoA***: long chain fatty acid CoA ligase; ***AD*****:** acyl dehydratase; ***FM***: FAD-binding monoxigenase; ***DH***: iron-containing alcohol dehydrogenase; ***dxnF***:

hydroxyquinol 1,2-dioxygenase (EC 1.13.11.37). **VII) homoprotocatechuate pathway**: ***xylE, hsaG, hsaF*** are previously described; ***CHDH***: 5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase (EC 1.2.1.60); **SCDH**: putative short chain dehydrogenase; ***tau***: 4-oxalocrotonate tautomerase; ***nit***: NADH-FMN oxidoreductase-nitrilotriacetate monooxygenase component B (EC 1.14.13.-); ***hpa***: 4-hydroxyphenylacetate 3-monooxygenase (EC 1.14.13.3). **VIII) A central pathway** with an unknown substrate described in *R. jostii* RHA1: ***duf1486****:* protein of unknown function DUF1486 (probable NADH dehydrogenase/NAD(P)H nitroreductase); ***acDH***: acyl-CoA dehydrogenase, type 2, C-terminal domain; ***DBPS***:3,4-dihydroxy-2-butanone 4-phosphate synthase (EC 4.1.99.12)/GTP cyclohydrolase II (EC 3.5.4.25); ***ox****:* NADH-FMN oxidoreductase; ***HP****:* hypothetical protein; ***DHBDII***: biphenyl-2,3-diol-1,2-dioxygenase II (EC 1.13.11.39) (2,3-dihydroxybiphenyl dioxygenase II); ***hpcE***: possible fumarylacetoacetate hydrolase; ***hyd*****:** FAD-binding monooxygenase (PheA/TfdB family), conserved hypothetical hydroxylase, similar to 2,4-dichlorophenol 6-monooxygenase; ***syn***: acetoacetyl-CoA synthetase (EC 6.2.1.16); ***asnC***: transcriptional regulator (AsnC family); ***pyrDH***: pyruvate dehydrogenase E1 component (EC 1.2.4.1). All RHA1 genes have a "RHA1_" as prefix to each gene number.

III. **The 2-hydroxypentadienoate pathway (Hpd)**. It transforms 2-hydroxypentadienoates to acetyl-CoA and pyruvate through successive actions of a hydratase, an aldolase, and a dehydrogenase (Figure 4. III) (van der Geize, Yam *et al.* 2007).

IV. **The gentisate pathway**. It converts gentisate to pyruvate and fumarate and it is encoded by the *genABC* genes (Figure 4. IV) (Suemori, Nakajima *et al.* 1995).

V. **The homogentisate pathway (Hmg).** Encoded by the *hmgABC* genes, this pathway involves the extradiol cleavage of homogentisate, followed by C-C bond hydrolysis to finally yield fumarate and acetoacetate (Figure 4 V) (Arias -Barrau, 2004).

VI. **The hydroxyquinol pathway**. It is a degradative pathway, related to intradiol-type cleavage of 4-hydroxysalicylate/hydroxyquinol to aceyl-CoA and succinyl-CoA (Figure 4 VI) (Armenguad 1999, Travkin, Solyanikova *et al.* 2006).

VII. **The homoprotocatechuate (3,4-dihydroxyphenylacetate)**. It is a pathway encoded by the *hpc* genes that are involved in the extradiol-type cleavage of homoprotocatechuate (Figure 4 VII).

VIII. **A central pathway**. In *R. jostii* RHA1, it has been found another pathway comprising: a hydroxylase, an extradiol dioxygenase and a hydrolase, although its substrate has not been identified yet (Figure 4 VIII) (McLeod, Warren *et al.* 2006; Yam, van der Geize *et al.* 2010).

Not all these 8 pathways are found in all *Rhodococci*, for instance, the genes encoding the gentisate, homoprotocatechuate, and the pathway VIII found in RHA1 are absent in at least two *R. erythropolis* strains, PR4 and SK121 (Yam, van der Geize *et al.* 2010).

R. ruber draft genome contains 7 out of 8 central aromatic pathways (only the Paa pathway is absent) (Figure 3A). In fact, most of the smallest *Rhodococcus* genomes lack this pathway (*e.g. R. equi*, *R. aetherivorans* or *R. pyridinivorans* among them) while is always present in larger *Rhodococcus* genomes.

Some of the genes detected in the central pathways are redundant in the genome. For instance, the cluster *pcaJ-pcaI* that encode a succinyl-CoA: 3-ketoacid-coenzyme A transferase (EC 2.8.3.5) are present in 3 copies in the *R. ruber* genome: contig NZ_ANGC02000003.1 (D092_RS10690-RS10695), contig NZ_ANGC02000026.1 (D092_RS24595-RS24600) and contig NZ_ANGC02000008.1 (D092_RS18665-RS18670). This activity is involved in the catechol and protocatechuate branch of β-ketoadipate pathway and in the leucine degradation metabolism. The muconate cycloisomerase (EC 5.5.1.1) that it is related to the catechol branch of β-ketoadipate pathway and in the chlorobenzoate degradation, is also present in 3 copies: contig 8 (D092_RS18510), contig NZ_ANGC02000001.1 (D092_RS03740) and contig NZ_ANGC02000004.1 (D092_RS12240). The 2-hydroxypentanodienoate pathway consists of three enzymes in tandem: a hydratase, an aldolase and a dehydrogenase. There are 2 copies of this tandem in the *R. ruber* genome (contig NZ_ANGC02000004.1; Figure 4 III) related to degradation of phenols, cresols or catechol.

2.6. Peripheral pathways

The peripheral pathways vary from one *Rhodococcus* to another due to the genomic size, plasmid content and the redundancy of genes. Some of the aromatic compounds that could serve as substrate to *R. ruber* and the genes encoding the putative necessary enzymatic activities are described below. On the other hand, other clusters could not be found such as the *bph* cluster for biphenyl catabolism, (Goncalves, Hara *et al.* 2006) and the *pad* cluster for phthalate catabolism (Patrauchan, Florizone *et al.* 2005).

A. Benzoate degradation

Rhodococcus jostii strain RHA1 catabolizes the benzoate via cathecol pathway through a ring-hydroxylating oxygenase (Patrauchan, Florizone *et al.* 2005). The cathecol and the protocatechuate branches of the β-ketoadipate pathway converge at the β-ketoadipate enol-lactone in this strain. All these clusters (*ben*, *cat* and *pca)* are present in *R. ruber* Chol-4 and are closely located in a way similar to *R. jostii* RHA1 (Figure 4 and 5A) (Patrauchan, Florizone *et al.* 2005).

B. Isopropylbenzene (cumene) degradation

Degradation of isopropylbenzene lies on the cluster *ipbA1A2A3A4C* in *Rhodococcus erythropolis* BD2. These genes encode a reductase (*ipbA4*), a ferredoxin (*ipbA3*), a dioxygenase (*ipbA1A2*) and a 3-isopropylcatechol-2,3-dioxygenase (*ipbC*) (Kesseler, Dabbs *et al.* 1996). The isopropylbenzene pahway is found within plasmids in the *Rhodococcus* genomes: the linear plasmid, pBD2 (208 to 212 kb) from *Rhodococcus erythropolis* BD2 (Dabrock, Kesseler *et al.* 1994), or the plasmid pRHL1 from *Rhodococcus jostii* RHA1 (NCBI: NC_008269). *R. ruber* contains this cluster in contig NZ_ANGC02000001.1 and a partial cluster in contig NZ_ANGC02000021.1 (Figure 5B). The *ipb* operon transforms isopropylbenzene to isobutyrate, pyruvate, and acetyl-coenzyme A in *Pseudomonas putida* RE204 (Kesseler, Dabbs *et al.* 1996; Eaton, Selifonova *et al.* 1998).

C. Steroids and cholic acid degradation

Rhodococci are so widely known as good steroid eaters (Yam and Okamoto 2011) that they could be defined as steroid-consumer strains by excellence. The cholesterol pathway has been extensively studied. Part of its complexity is due to alternative pathways and the diversity of the enzymes that are involved. Just a couple of examples: the first step could be performed by a 3β-hydroxysteroid dehydrogenase or a cholesterol oxidase, and the side-chain degradation could occur at different steps of the pathway depending on the strain (Yam, van der Geize *et al.* 2010).

Up to 4 clusters of genes encoding distinct steroid catabolic pathways have been reported in RHA1 (McLeod, Warren *et al.* 2006). The cholesterol catabolism proceeds via oxidative opening of ring B with concomitant aromatization of ring A (Yam and Okamoto 2011).

The ability to grow in different steroids (such as cholesterol, cholestenone, testosterone, AD or ADD) and the role of ketosteroid dehydrogenases (KstDs) and cholesterol oxydase in *R. ruber* has been previously reported (Fernández de las Heras, García Fernández *et al.* 2009; Fernández de las Heras, van der Geize *et al.* 2012; Fernández de las Heras, Perera *et al.* 2014). In Figure 5C some of the steroid clusters contained in the *R. ruber* genome are depicted. The steroid degradation genes are described to occur within large gene clusters (Bergstrand, Cardenas *et al.* 2016) and this seems also to happen in *R. ruber*.

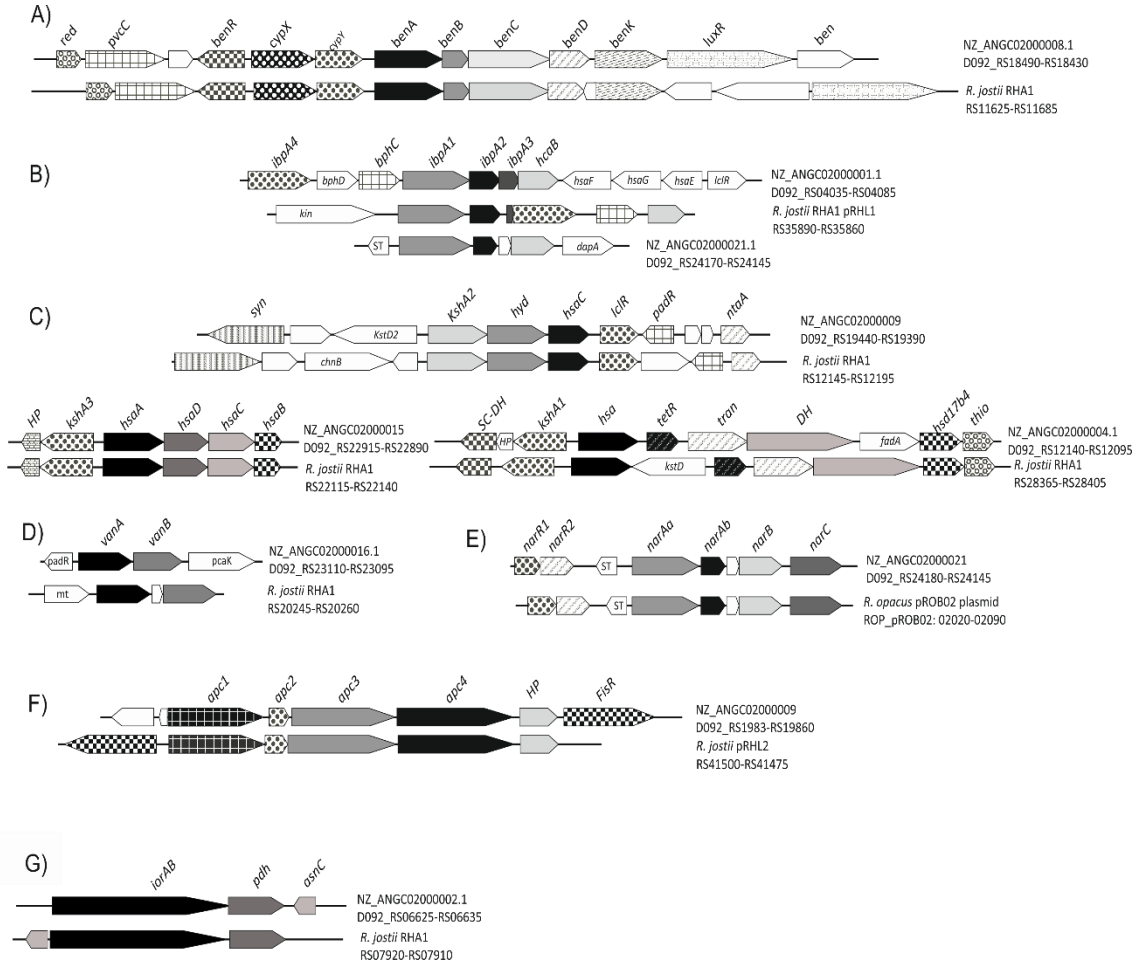


Figure 5. Peripherical routes in *R. ruber*: Abbreviations: **A) Benzoate degradation**: ***red***: flavin reductase; ***pvcC***: pyoverdin chromophore biosynthetic protein; ***benR***: transcriptional regulator (AraC family); ***cypX***: cytochrome P450 monooxygenase; ***cypY***: putative phenol hydroxylase; ***benA***: benzoate 1,2-dioxygenase α subunit (EC 1.14.12.10); ***benB***: benzoate 1,2-dioxygenase β subunit (EC 1.14.12.10); ***benC***: benzoate dioxygenase, ferredoxin reductase component/1,2-dihydroxycyclohexa-3,5-diene-1-carboxylate dehydrogenase (EC 1.3.1.25); ***benD***: 1,2-dihydroxycyclohexa-3,5-diene-1-carboxylate dehydrogenase (EC 1.3.1.25); ***benK***: benzoate MFS transporter; ***luxR***: transcriptional regulator (luxR family) putative.; ***ben***: benzoate transport protein. **B) Isopropylbenzene pathway**: ***ipbA4***: ferredoxin reductase; ***bphD:*** 2-hydroxy-6-oxo-2,4-heptadienoate hydrolase; ***bphC***: 2,3-dihydroxybiphenyl 1,2-dioxygenase; ***ipbA1***: isopropylbenzene 2,3-dioxygenase or IPB-dioxygenase, ISP large subunit; ***ipbA2***: IPB-dioxygenase (ISP small subunit); ***ipbA3***: IPB-dioxygenase ferredoxin; ***hcaB***: hydroxybenzaldehyde dehydrogenase***; hsaF, hsaG; hsaE***: previously described (Figure 4); ***lclR***: transcriptional regulator (IclR family); ***ST***: sterol-binding domain protein; ***Kin***: sensor kinase; ***dapA***: 4-hydroxy-tetrahydrodipicolinate synthase (EC 4.3.3.7). **C) Steroids pathway**: ***syn***: non-ribosomal peptide synthetase; ***kstD***: 3-oxosteroid 1-dehydrogenase; ***kshA***: ketosteroid-9-α-hydroxylase, oxygenase(EC:1.14.13.); ***hyd***: hydroxylase; ***hsaC***: 2,3-dihydroxybiphenyl 1,2-dioxygenase; ***lclR***: transcriptional regulator (IclR family); ***padR***: transcriptional regulator (PadR family); ***ntaA***: nitrilotriacetate monooxygenase component B (EC 1.14.13.-); ***HP***: hypothetical protein; ***hsaA***: flavin-dependent monooxygenase; ***hsaD***: 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (EC 3.7.1.-); ***hsaC***: iron-dependent extradiol dioxygenase; ***hsaB***: flavin-dependent monooxygenase reductase subunit; ***SC-DH***: short-chain dehydrogenase; ***hsa***: monooxygenase; ***tetR***: probable transcriptional regulator (TetR family); ***tran***: acetyl-CoA acetyltransferase; ***DH***: acyl-CoA dehydrogenase; ***fadA***: 3-ketoacyl-CoA thiolase (EC 2.3.1.16); ***hsd17b4***: 3-α,7-α,12-α-trihydroxy-5-β-cholest-24-enoyl-CoA hydratase; ***thio***: thioesterase; ***HP***: hypothetical protein; ***chnB***: cyclohexanone monooxygenase. **D) Vanillate (4-hydroxy-3-methoxybenzoate)**: ***padR***: transcriptional regulator (PadR family); ***vanA***: vanillate *o*-demethylase oxygenase subunit (EC 1.14.13.82), flavodoxin reductases (ferredoxin-NADPH reductases) family 1; ***vanB***: vanillate *o*-demethylase oxidoreductase (EC 1.14.13.-); ***pcaK***: 4-hydroxybenzoate

transporter; mt:methyltransferase. **E) Naphtalen**. **(*nar* genes*: R. opacus* plasmid pROB02:AP011117)**: ***narR1*** and ***narR2***: putative naphthalene degradation regulatory protein; ***narAa***: nidA, naphthalene dioxygenase large subunit; ***narAb***: (nidB) naphthalene dioxygenase small subunit; ***narB***: (nidC) 1,2-dihydro-1,2-dihydroxynaphthalene dehydrogenase; ***narC***: (nidD) putative aldolase NarC. **F) Acetophenone carboxylase (anaerobic)**: ***apc1-4***: acetophenone carboxylase subunits; ***HP***: hypothetical protein; ***fisR***: transcriptional regulato (Fis family). **G) Aminoacid**: ***iorAB***: indolepyruvate ferredoxin oxidoreductase (α and β subunits); ***pdh***: glutamate/leucine/phenylalanine/valine dehydrogenase; ***asnC*****:** transcriptional regulator (AsnC family). All RHA1 genes have a "RHA1_" as prefix to each gene number.

The Mce system is an ATP-binding cassette transporter comprising more than eight distinct proteins that are present in actinobacteria. The number of Mce systems could vary within bacteria: from 4 in *Mycobacterium tuberculosis* H37Rv to 6 in *M. smegmatis*. The Mce4 system of *Rhodococcus jostii* RHA1 or *Mycobacterium smegmatis* has been proved to be an active uptake system that requires ATP to transport steroids such as cholesterol, 5-α-cholestanol, 5-α-cholestanone or β-sitosterol (Mohn, van der Geize *et al.* 2008; Klepp, Forrellad *et al.* 2012). The other Mce operons could be involved in the cell envelope structure maintenance of its biosynthesis processes (Klepp, Forrellad *et al.* 2012). **T**here are at least 3 *mc*e systems in *R. ruber* (Figure 6). The one that keeps a bigger similarity with the mce4 of RHA1 is the system of Contig NZ_ANGC02000015.1 (Figure 6).

Two transporters, the ABC-transporter *CamABCD* (ro04888 to ro04885), and a gene encoding an MFS transporter, *CamM* (ro05792), are associated with cholate catabolism in *Rhodococcus jostii* RHA1 (Swain, Casabon *et al.* 2012). However, contrary to the *mce* cluster, there are no ORFs similar detected to these transporters in the *R. ruber* draft genome suggesting either other kind of transports acting or a different range of metabolic byproducts formed in the process. It has been proposed recently that the cholate pathway genes were derived from gene duplication of a single ancestral 9,10-seco-steroid degradation pathway and that they are conserved and part of the core genome of the genus *Rhodococcus* (Bergstrand, Cardenas *et al.* 2016). The cholate catabolic cluster confirmed in RHA1 (Mohn, Wilbrink *et al.* 2012) is also located in the *R. ruber* genome (Figure 4C) and growth on this substrate was also confirmed in Chol-4. However, steroid transport could have evolved for other ways and therefore, it is possible that not all *Rhodoccus* share the same mechanisms for it.

D. Vanillate degradation (4-hydroxy-3-methoxybenzoate)

Vanillate is a lignin-derived methoxylated monocyclic aromatic compound. A cluster similar to the vanillate genes of *Pseudomonas* sp. strain HR199, the gene loci *vanA* and *vanB* is found in *R. ruber* genome (GenBank::Y11521, Figure 5D). The vanillate catabolism proceeds via

protocatechuate in *Comamonas testosteroni* strain BR6020 and in *Pseudomonas* sp. strain HR199 (Priefert, Rabenhorst *et al.* 1997; Providenti, O'Brien *et al.* 2006).

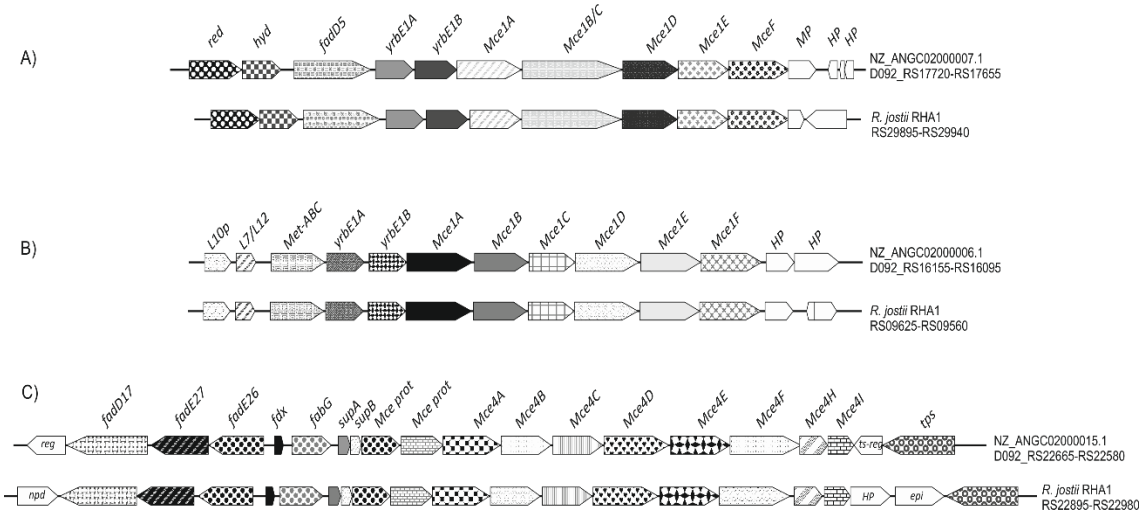


Figure 6: *Mce* systems in *R. ruber*. Abbreviations: **A)** ***red***: enoyl-[acyl-carrier-protein] reductase [FMN] (EC 1.3.1.9); ***hyd*****:** enoyl-CoA hydratase (EC 4.2.1.17); ***fadD5***: long-chain fatty-acid-CoA ligase (EC 6.2.1.3), *Mycobacterial* subgroup FadD5; ***yrbE1A***: conserved hypothetical integral membrane protein YrbE1A (ABC-transporter permease); ***yrbE1B***: conserved hypothetical integral membrane protein YrbE1B (ABC-transporter permease); ***mce1A-D***: MCE-family protein MceA-D; ***mce1e***: MCE-family lipoprotein LprK (MCE-family lipoprotein Mce1e); ***mceF***: MCE family protein of Mce F Subgroup; ***MP***: membrane protein; ***HP***: hypothetical protein; ***pep-ABC***: peptide ABC-transporter permease. **B)** ***L10p***: LSU ribosomal protein L10p (P0); ***L7/L12***: LSU ribosomal protein L7/L12 (P1/P2); ***met-ABC***: methionine ABC-transporter ATP-binding protein (*npd*:2-nitropropane dioxygenase, NPD). **C)** ***reg***: possible transcriptional regulatory protein; ***fadD17***: long-chain fatty-acid-CoA ligase (EC 6.2.1.3) (*Mycobacterial* subgroup FadD17); ***fadE27***: butyryl-CoA dehydrogenase (EC 1.3.99.2); ***fadE26***: acyl-CoA dehydrogenase (EC 1.3.8.1) (*Mycobacterial* subgroup FadE26); ***fdx***: ferredoxin; ***fabG***: 3-ketoacyl-ACP reductase (also called hsd4A); ***supA*** and ***supB***: ABC-transporter permease; ***ts-reg****:* two-component system response regulator; ***tps***: α,α-trehalose-phosphate synthase [UDP-forming] (EC 2.4.1.15); ***epi***: epimerase, dihydroflavonol-4-reductase (EC 1.1.1.219).

E. Naphtalene

Rhodococcus sp. strain NCIMB 12038 and *Rhodococcus opacus* R7 degrade naphthalene via the *nar* cluster encoding a gentisate 1,2-dioxygenase (converts gentisate to maleylpyruvate), a mycothiol-dependent maleylpyruvate isomerase (catalyzes the isomerization of maleylpyruvate to fumarylpyruvate) and a fumarylpyruvate hydrolase (hydrolyzes fumarylpyruvate to fumarate and pyruvate) (Di Gennaro, Terreni *et al.* 2010; Liu, Xu *et al.* 2011). The gentisate degradation pathway is shared by both naphthalene and 3-hydroxybenzoate catabolism. The *nar* gene cluster presents a diverse genetic organization with even different kind of regulators within *Rhodococcus* strains (Di Gennaro, Terreni *et al.* 2010).The *nar* cluster found in *R. ruber* Chol-4 (Figure 5E) is similar to the cluster found in the plasmid pROB02 of *Rhodococcus opacus* B4

(NC_012521). Two regulatory genes, *narR1* belonging to the GntR family, and NarR2 of the NtrC family of enhancer-binding proteins are found.

Rhodococcus sp. strain TFB is reported to use three coordinated set of genes to degrade naphthalene via salicylaldehyde and gentisate to yield fumarate and pyruvate, among them the dioxygenase *thnA1234* cluster that could correspond to the *ipb1234* cluster found in the *R. ruber* genome (Tomas-Gallardo, Gomez-Alvarez *et al.* 2014). Following this scheme, the isopropylbenzene cluster (Figure 5B) could be also involved in the *R. ruber* naphtalen degradation.

F. Acetophenone carboxylase (anaerobic)

A putative anaerobic degradation cluster that encodes an acetophenone carboxylase has been also found in the genome of *R. ruber* in contig NZ_ANGC02000009.1 (Figure 5F).The acetophenone carboxylase is involved in the anaerobic degradation of ethylbenzene in the betaproteobacterium *Aromatoleum aromaticum:* ethylbenzene is converted by anaerobic oxidation to acetophenone via (*S*)-1-phenylethanol and the subsequent carboxylation of acetophenone to benzoylacetate is catalyzed by the acetophenone carboxylase (Jobst, Schuhle *et al.* 2010). This cluster is also present in a plasmid pRHL2 in *R. jostii* RHA1 (Figure 5 F) although the biological role that this enzyme could have in *Rhodococcus* strains is unknown. Recently, a cross-regulation between anaerobic and aerobic pathways for the catabolism of aromatic compounds has been shown in the β-proteobacterium *Azoarcus sp*., and it may reflect a biological strategy to increase the cell fitness in organisms that survive in environments subject to changing oxygen concentrations (Valderrama, Durante-Rodriguez *et al.* 2012).

G. Amino acids degradation

Pseudomonas putida metabolizes L-phenylalanine, 3-hydroxyphenylacetate and L-tyrosine through a peripheral pathway with the formation of homogentisate as the central intermediate (Arias-Barrau, Olivera *et al.* 2004). Phenylalanine dehydrogenase (Pdh, EC 1.4.1.20) catalyzes the reversible pyridine nucleotide-dependent oxidative deamination of L-phenylalanine to yield phenylpyruvate and ammonia (Brunhuber, Thoden *et al.* 2000). This enzyme is also found in the *R. ruber* genome (Figure 5G).

2.7. Growth experiments

In order to determine the growth capabilities of *Rhodococcus ruber* strain Chol-4, we have analyzed its ability to use several compounds as sole energy and carbon sources. The preliminary results have shown that this strain is able to grow in minimal medium supplemented with a single compound as source of carbon and energy such as: isopropanol, 1,3-butanediol, 2,3-butanediol (all of them as a vapor), 1 mg/mL of naphthalene in powder, 10 mM sodium benzoate, 2 mM phenol, 5 mM L-tryptophan, 4 mM vanillic acid, 4 mM gentisate, 5 mM homogentisate, 2 mM catechol, 2.2 mM cholic acid, 2 mM dehydroepiandrosterone (DHEA) and 10 mM protocathecuate (pca). On the other hand, this strain was not able to grow in the majority of volatile compounds (BETEX: benzene, toluene, ethylbenzene and xylene isomers, phenylacetic acid, indane, tetralin and styrene) as well as in other compounds such as biphenyl, phthalates, 2-aminobenzoate, salicylic acid (2-hydroxybenzoic acid), 1,2,4-benzenetriol (hydroxyquinol) or L-tyrosine. These data is in accordance with the theoretical clusters found within the genome of *Rhodococcus ruber*: there were clusters related with the catabolism of gentisate, catechol, pca, naphthalene, phenol, cholic acid among others (Figure 3 and 4). But there are some exceptions: *R. ruber* could not grow on hydroxyquinol even though it seems to contain the pathway VI (see Figure 3 VI). It could also not grow on salicylate, although up to two putative salicylate hydroxylase (EC 1.14.13.1) are found in the genome (contig NZ_ANGC02000001.1, D092_RS04015 and contig NZ_ANGC02000006.1, D092_RS16585). It is possible that other factor such as transport of the compound could be affecting. In other *Rhodococcus* strains, degradation of many xenobiotic compounds has been acquired by gene transfer: haloalkanes, alkenes, biphenyl or naphthalene among others (Larkin, Kulakov *et al.* 2010).

Table 1. Growth data of *Rhodococcus ruber* strain Chol-4

Carbon source	Doubling time (hours)	Maximum OD_{260nm}
2.2 mM Cholic acid in CD	1.8 ± 0.1	2.6 ± 0.2
10 mM Sodium benzoate	2.2 ± 0.2	2.5 ± 0.1
2 mM DHEA in CD	7.0 ± 1.5	1.3 ± 0.1
8 mM Naphtalen	15.9 ± 3.3	0.2 ± 0.1
4 mM Gentisate	2.8 ± 0.4	1.0 ± 0.2

These data are the result of unless triplicate experiments. CD: 16.5 mM cyclodextrins

The growth curves of some of the compounds in which *R. ruber* could grow were determined here (Figure 7, Table 1). *R. ruber* was able to grow in minimal medium 457 supplemented with sodium benzoate 10 mM as sole organic substrate.

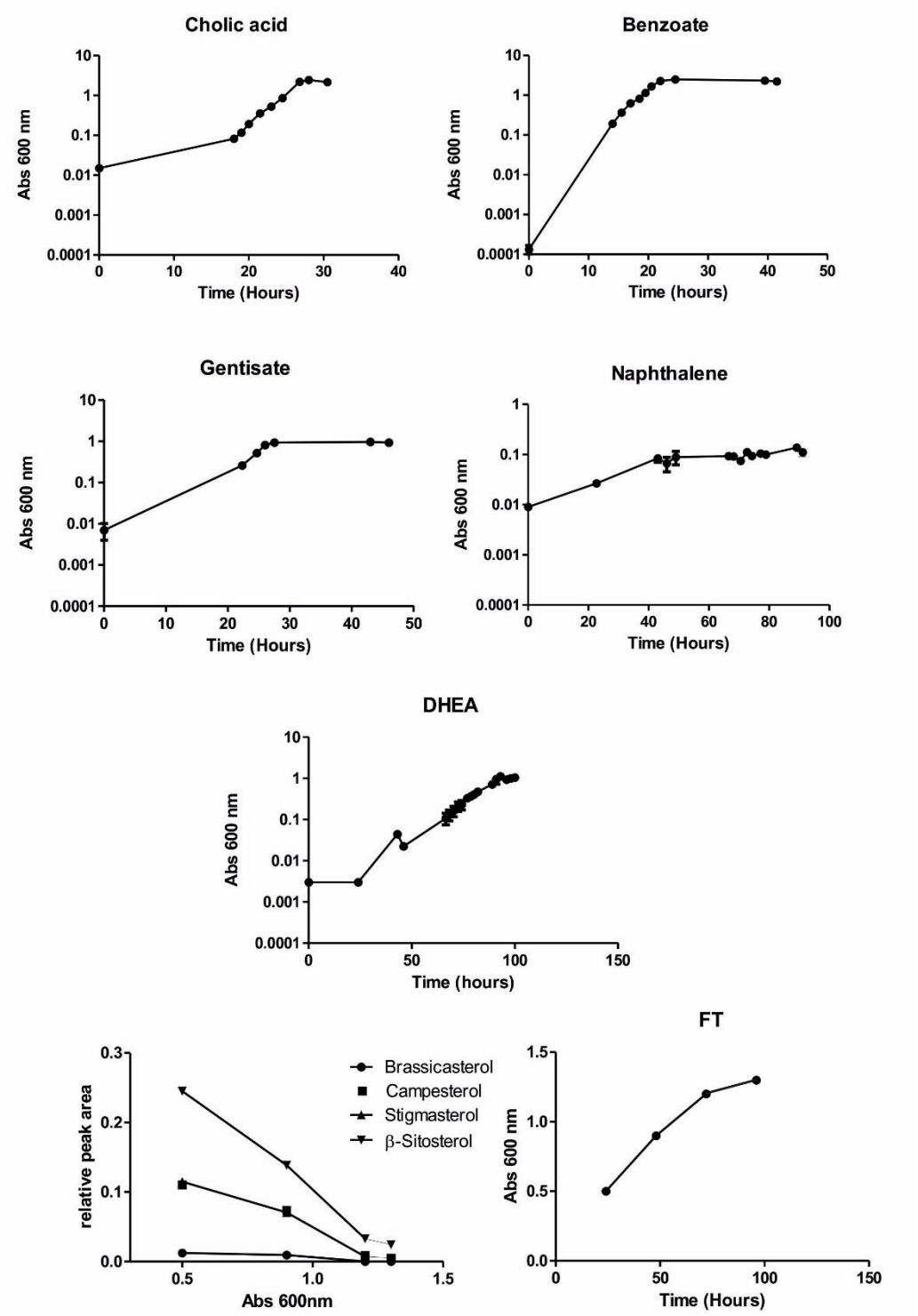


Figure 7. Growth curves of *R. ruber* WT in different compounds. *R. ruber* was grown in minimal medium 457 supplemented with: 10 mM benzoate sodium, 4 mM gentisate, or 2.1 mM DHEA in the presence of 16.5 mM cyclodextrins. No growth when using only cyclodextrins was appreciated. *R. ruber* was also grown in minimal medium 457 supplemented with naphthalene in powder (1 mg/mL) and a mixture of phytosterols (brassicasterol, campesterol, stigmasterol and β-stosterol). Data of 3-4 independent experiments are depicted. In the case of phytosterols, the relative peak area is the ratio between the HPLC peak obtained for each phytosterol and the peak of pregnenolone used as internal control.

On the other hand, *R. ruber* also grows on catechol and protocatechuic acid, and therefore the catabolism of the benzoate could be via cathecol pathway through a ring-hydroxylating oxygenase as it has proposed for RHA1 (Patrauchan, Florizone *et al.* 2005).

In this work, the steroid knowledge about the degradation capabilities of *R. ruber* has been broaden to phytosterols (plant steroids), DHEA and cholic acid (a primary bile acid). *R. ruber* was able to grow in minimal medium 457 supplemented with a mix of phytosterols (plant sterols that include brassicasterol, campesterol, stigmasterol and β-sitosterol) and its consumption was followed by HPLC (Figure 7). *R. ruber* cells also grew in minimal medium 457 with cholic acid or DHEA as a sole organic source (Figure 7, Table 1).

R. ruber was able to grow in minimal medium 457 with naphthalene in powder (1 mg/mL) and also in minimal medium 457 supplemented with 4 mM gentisate as sole organic substrates (Figure 7 and Table 1). *R. ruber* was also grown in minimal medium 457 with 4 mM gentisate. As it has been stated before, there are two different pathways described for naphthalene catabolism in *Rhodococcus*: using the *nar* cluster or using the isopropylbenzene cluster but both via salicylate and gentisate. However, *R. ruber* could not grow on salicylate and therefore the pathway that this strain uses to catabolize naphthalene remains unclear.

2.8. Mutant construction

In order to experimentally check the functionality of some of the pathways found in *R. ruber*, two group of genes were deleted: protocatechuate 3,4-dioxygenase α chain (*pcaG*) and 3-carboxy-cis,cis-muconate cycloisomerase (*pcaB*) from the cluster related to the protocatechuic acid (pathway aromatic I, Figure 8A and 4I) and some naphthalene putative genes including the naphthalene dioxygenase (*nar* genes, Figure 8A and 5E).

Unmarked gene deletions were carried out in the *pca* and *nar* genes in *R. ruber* strain Chol-4 to verify the roles of these genes in the growth of protocatechuate and naphthalene, respectively. A scheme of the deletions introduced is depicted in Figure 8 A. Mutants were confirmed by PCR and growth experiments probed that *nar R. ruber* mutants were not able to grow on naphthalene; similarly, *pca R. ruber m*utants were not able to grow on protocatechuate (Figure 8 B).

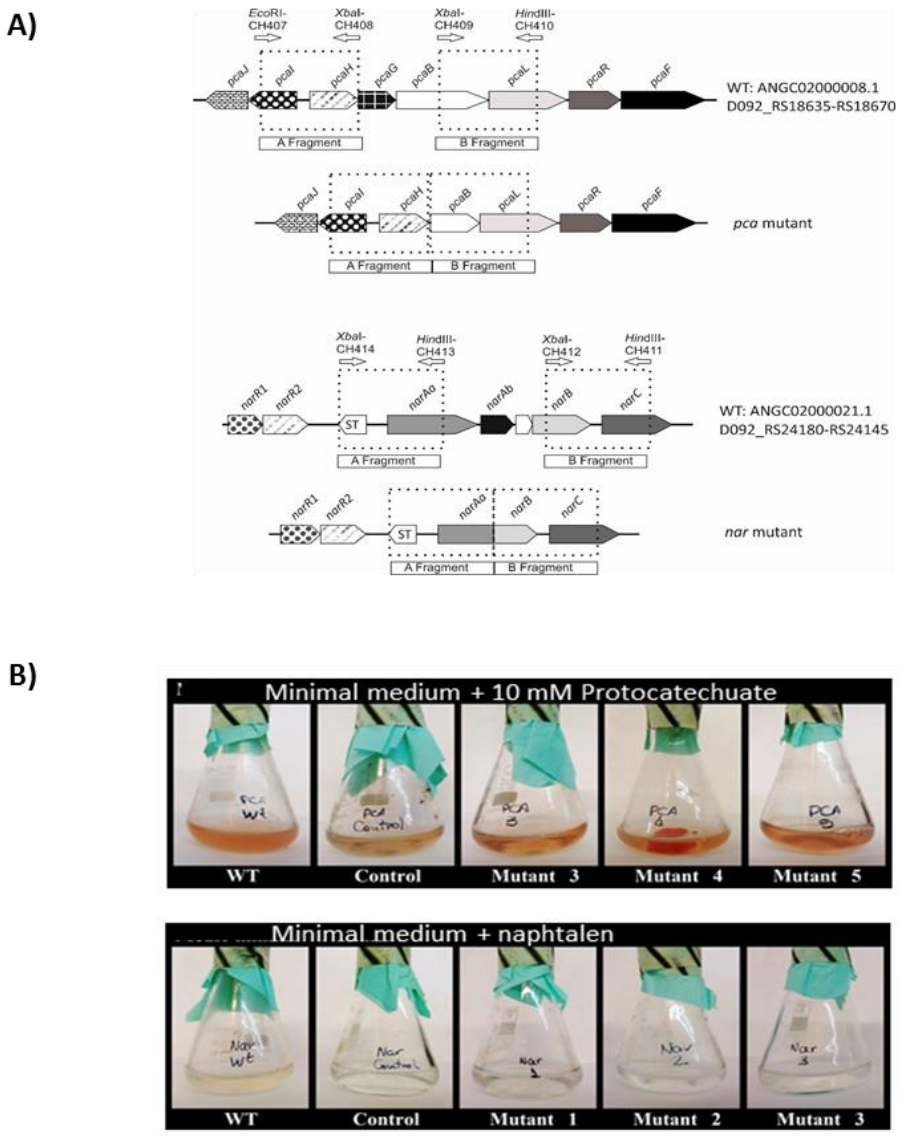


Figure 8. *Rhodococcus ruber* mutants. **A)** Scheme of *R. ruber* deletions of this work for *pca* mutant and *nar* mutants**. B)** Growth of *R. ruber* on minimal medium supplemented with 10 mM PCA or naphthalene in powder (1 mg/mL), respectively. WT: wild type; *pca* and *nar* mutants (1-3); control: growth in the absence of inoculum.

On the other hand, the growth of the *nar* and *pca R. ruber* mutants were also checked for different substrates (Table 2).The Δ*nar* mutant could grow in all substrates tested, but the Δ*pca* mutant failed to grow also in vanillate. Therefore, the growth in this compound is dependent also of the *pca* cluster (Figure 3A).

Table 2. Growth of *R. ruber* Chol-4 wild type and mutants on minimum medium with different substrates.

Strain/carbon source	Naphthalen 7 mM	Benzoate 10 mM	L-Tryp 5 mM	Catechol 2 mM	Homog 2 mM	Gentisate 4 mM	PCA 10 mM	Chol mM	Vanill 4 mM
Wild type	+	+	+	+	+	+	+	+	+
Δ*nar*	-	+	+	+	+	+	+	+	+
Δ*pca*	+	+	+	+	+	+	-	+	-

L-Tryp: L-Tryptophan, Homog: Homogentisate; PCA: protocatechuic acid; Chol: cholic acid; Vanill: vanillic acid

R. ruber is able to grow in minimal medium with steroids (*e.g.* cholesterol, phytosterols, DHEA), bile acids (cholic acid) or aromatic compounds (*e.g.* benzoate, naphthalene, gentisate) as the only source of carbon and energy. The results confirm that *R. ruber* strain Chol-4 has a high biotechnological interest due to its metabolic potential, for instance, the bacterial transformation of steroids to produce pharmaceutically active steroid drugs. Our findings also provided a valuable guide for future in depth molecular, microbiological or physiological studies of *R. ruber* and provide a basis for the use of this strain in the field of Biotechnology.

SUPLEMENTAL MATERIAL

Table S1. Anticodons encoded in the *R. ruber* Chol-4 genome.

Num. tRNAs	Amino acid	Cd.	Ant. Cd.	Contig	Start	End	L	Locus tag
1	Alanine (A)	GCC	tRNA-Ala-GGC	NZ_ANGC02000002.1	187388	187460	73	D092_RS05650
2	Alanine (A)	GCC	tRNA-Ala-GGC	NZ_ANGC02000002.1	187537	187609	73	D092_RS05655
3	Alanine (A)	GCG	tRNA-Ala-CGC	NZ_ANGC02000001.1	422777	422852	73	D092_RS02040
4	Alanine (A)	GCA	tRNA-Ala-TGC	NZ_ANGC02000005.1	411740	411815	73	D092_RS15550
	Alanine (A)	GCT	-					
5	Arginine (R)	CGG	tRNA-Arg-CCG	NZ_ANGC02000022.1	2562	2487	73	D092_RS24260
6	Arginine (R)	CGT	tRNA-Arg-ACG	NZ_ANGC02000023.1	22641	22713	73	D092_RS24510
7	Arginine (R)	AGA	tRNA-Arg-TCT	NZ_ANGC02000003.1	32114	32038	74	D092_RS08525
8	Arginine (R)	AGG	tRNA-Arg-CCT	NZ_ANGC02000001.1	338994	338922	73	D092_RS01630
	Arginine (R)	CGC	-	-	-	-	-	
	Arginine (R)	CGA	-	-	-	-	-	
9	Asparagine (N)	AAC	tRNA-Asn-GTT	NZ_ANGC02000002.1	339525	339597	73	D092_RS06395
	Asparagine (N)	AAT	-	-	-	-	-	
10	Aspartic acid (D)	GAC	tRNA-Asp-GTC	NZ_ANGC02000001.1	170705	170778	74	D092_RS00775
11	Aspartic acid (D)	GAC	tRNA-Asp-GTC	NZ_ANGC02000001.1	170964	171037	74	D092_RS00785
	Aspartic 7acid (D)	GAT	-	-	-	-	-	
12	Cysteine (C)	TGC	tRNA-Cys-GCA	NZ_ANGC02000003.1	400974	400904	71	D092_RS10200
	Cysteine (C)	TGT	-	-	-	-	-	
13	Glutamic acid (E)	GAG	tRNA-Glu-CTC	NZ_ANGC02000002.1	273884	273959	73	D092_RS06065
14	Glutamic acid (E)	GAG	tRNA-Glu-CTC	NZ_ANGC02000002.1	274019	274094	73	D092_RS06070
15	Glutamic acid (E)	GAA	tRNA-Glu-TTC	NZ_ANGC02000001.1	170575	170647	73	D092_RS00770
16	Glutamine (Q)	CAG	tRNA-Gln-CTG	NZ_ANGC02000002.1	273720	273791	72	D092_RS06060
17	Glutamine (Q)	CAA	tRNA-Gln-TTG	NZ_ANGC02000001.1	487465	487394	72	D092_RS02330
18	Glycine (G)	GGG	tRNA-Gly-CCC	NZ_ANGC02000004.1	21346	21416	71	D092_RS1178
19	Glycine (G)	GGA	tRNA-Gly-TCC	NZ_ANGC02000003.1	74274	74201	71	D092_RS08710
20	Glycine (G)	GGC	tRNA-Gly-GCC	NZ_ANGC02000003.1	400548	400473	73	D092_RS10180
21	Glycine (G)	GGC	tRNA-Gly-GCC	NZ_ANGC02000003.1	400766	400691	73	D092_RS10190
22	Glycine (G)	GGC	tRNA-Gly-GCC	NZ_ANGC02000003.1	401072	400997	73	D092_RS10205
	Glycine (G)	GGT	-	-	-	-	-	
23	Histidine (H)	CAC	tRNA-His-GTG	NZ_ANGC02000003.1	4850	4775	73	D092_RS08395
	Histidine (H)	CAT	-	-	-	-	-	
24	Isoleucine (I)	ATC	tRNA-Ile-GAT	NZ_ANGC02000005.1	411528	411601	74	D092_RS15545
	Isoleucine (I)	ATT	-	-	-	-	-	
	Isoleucine (I)	ATA	-	-	-	-	-	
25	Leucine (L)	CTG	tRNA-Leu-CAG	NZ_ANGC02000014.1	28859	28944	83	D092_RS22275
26	Leucine (L)	TTA	tRNA-Leu-TAA	NZ_ANGC02000001.1	530162	530238	74	D092_RS02525

Num. tRNAs	Amino acid	Cd.	Ant. Cd.	Contig	Start	End	L	Locus tag
27	Leucine (L)	CTA	tRNA-Leu-TAG	NZ_ANGC02000018.1	31076	30992	82	D092_RS23655
28	Leucine (L)	CTC	tRNA-Leu-GAG	NZ_ANGC02000003.1	625844	625759	86	D092_RS11260
29	Leucine (L)	CTC	tRNA-Leu-GAG	NZ_ANGC02000003.1	626122	626037	86	D092_RS11265
30	Leucine (L)	TTG	tRNA-Leu-CAA	NZ_ANGC02000002.1	611221	611145	74	D092_RS07605
	Leucine (L)	CTT	-	-	-	-	-	
31	Lysine (K)	AAA	tRNA-Lys-TTT	NZ_ANGC02000001.1	169593	169521	73	D092_RS00760
32	Lysine (K)	AAG	tRNA-Lys-CTT	NZ_ANGC02000003.1	1414	1342	73	D092_RS08380
33	Methionine (M)	ATG	tRNA-Met-CAT	NZ_ANGC02000002.1	170324	170400	74	D092_RS05550
34	Methionine (M)	ATG	tRNA-Met-CAT	NZ_ANGC02000002.1	339811	339887	74	D092_RS06400
35	Methionine (M)	ATG	tRNA-Met-CAT	NZ_ANGC02000006.1	139552	139477	73	D092_RS16200
36	Methionine (M)	ATG	tRNA-Met-CAT	NZ_ANGC02000008.1	69116	69040	74	D092_RS18695
37	Phenylalanine (F)	TTC	tRNA-Phe-GAA	NZ_ANGC02000001.1	170816	170892	74	D092_RS00780
	Phenylalanine (F)	TTT	-	-	-	-	-	
38	Proline (P)	CCG	tRNA-Pro-CGG	NZ_ANGC02000007.1	36608	36681	74	D092_RS17315
39	Proline (P)	CCA	tRNA-Pro-TGG	NZ_ANGC02000003.1	74412	74488	74	D092_RS08715
40	Proline (P)	CCC	tRNA-Pro-GGG	NZ_ANGC02000002.1	737857	737930	74	D092_RS08120
	Proline (P)	CCT	-	-	-	-	-	
41	Serine (S)	AGC	tRNA-Ser-GCT	NZ_ANGC02000023.1	22478	22569	89	D092_RS24505
42	Serine (S)	TCA	tRNA-Ser-TGA	NZ_ANGC02000023.1	16180	16267	85	D092_RS24475
43	Serine (S)	TCG	tRNA-Ser-CGA	NZ_ANGC02000019.1	6341	6428	88	D092_RS23745
44	Serine (S)	TCC	tRNA-Ser-GGA	NZ_ANGC02000007.1	248085	248172	85	D092_RS18320
	Serine (S)	AGT	-	-	-	-	-	
	Serine (S)	TCT	-	-	-	-	-	
45	Threonine (T)	ACA	tRNA-Thr-TGT	NZ_ANGC02000001.1	5702	5776	72	D092_RS00030
46	Threonine (T)	ACC	tRNA-Thr-GGT	NZ_ANGC02000006.1	139669	139597	73	D092_RS16205
47	Threonine (T)	ACG	tRNA-Thr-CGT	NZ_ANGC02000007.1	1059	984	73	D092_RS17135
	Threonine (T)	ACT	-	-	-	-	-	
48	Tryptophan (W)	TGG	tRNA-Trp-CCA	NZ_ANGC02000006.1	138001	137926	73	D092_RS16180
49	Tyrosine (Y)	TAC	tRNA-Tyr-GTA	NZ_ANGC02000006.1	139981	139896	83	D092_RS16210
	Tyrosine (Y)	TAT	-	-	-	-	-	
50	Valine (V)	GTA	tRNA-Val-TAC	NZ_ANGC02000002.1	358060	358134	72	D092_RS06490
51	Valine (V)	GTC	tRNA-Val-GAC	NZ_ANGC02000003.1	400683	400609	72	D092_RS10185
52	Valine (V)	GTC	tRNA-Val-GAC	NZ_ANGC02000003.1	400902	400828	72	D092_RS10195
53	Valine (V)	GTG	tRNA-Val-CAC	NZ_ANGC02000003.1	401383	401454	72	D092_RS10210
	Valine (V)	GTT	-	-	-	-	-	
	Stop codon	TGA	-					
	Stop codon	TAG	-					
	Stop codon	TAA	-					

Cd.: codon; **Ant. Cd.**: anticodon; **L**: length

Table S2. List of mobile elements find in the *R. ruber* Chol-4 genome.

Contig	Locus	Location Complement	Function	Max identity; Size; % identity/ % Similarity
NZ_ANGC02000001.1	D092_RS04160	894293..89596	Plasmid pria4b orf-3 family protein (phage integrase)	Plasmid pria4b ORF-3 family protein [*Rhodococcus rhodochrous* ATCC 21198] ETT25674.1;593aa 479/555(86%); 509/555(91%)
	D092_RS04185	905171..905756	Transposase (pseudogen)	
	D092_RS04190	905769..906277	Transposase (pseudogen	
	D092_RS04720	1021617..1022753	Phage integrase , site-specific tyrosine recombinase pham107	Integrase [*Gordonia aichiensis*] WP_005179717; 369aa 237/367(65%); 278/367(75%)
	D092_RS04730	1023799..1024926	Hypothetical protein;DNA helicase, phage-associated #pham380; repa	Hypothetical protein [*Nocardia sp*. NRRL WC-3656] WP_030515513.1; 333aa 196/322(61%); 241/322(74%)
NZ_ANGC02000002.1	D092_RS07855	667672..668442	ATP-binding protein/transposition helper	ATP-binding protein [*Mycobacterium kumamotonense*] WP_065289691.1; 255aa 232/253(92%); 244/253(96%)
	D092_RS07860	668439..669950	Mobile element protein	Transposase [*Mycobacterium sp.* YC-RL4] ANE81404.1; 505aa 409/501(82%); 442/501(88%)
	D092_RS08235	765763..766713	Putative integrase/recombinase	Recombinase [*Gordonia sp.* SGD-V-85] KSU51113.1; 345aa 315/316(99%); 316/316(100%)
	D092_RS08280	773291..776302	Dde transposase	DDE transposase [*Gordonia terrae*] WP_004018622.1; 1003aa 1002/1003(99%); 1003/1003(100%)
NZ_ANGC02000003.1	D092_RS09470	240258..241452	Transposase (pseudogen)	Transposase [*Pseudonocardia asaccharolytica*] WP_028932083.1; 138aa 127/138(92%); 132/138(95%)
	D092_RS09475	241478..241948	Mobile element protein: transposase	Transposase [*Nocardia elegans*] WP_063023642.1; 350aa 328/350(94%); 338/350(96%)
	D092_RS10710	506067..507119	3-keto-5-aminohexanoate cleavage protein//transposase	
	D092_RS11385	652545..653603	Integrase	Integrase [*Mycobacterium fortuitum*] WP_061264304.1; 377aa 174/348(50%); 213/348(61%)
NZ_ANGC02000004.1	D092_RS11825	31773..32825	3-keto-5-aminohexanoate cleavage protein//transposase	Transposase [*Nocardia salmonicida*] WP_062985089.1; 350aa 317/350(91%); 334/350(95%)
	D092_RS13055	299545..301230	Mobile element protein	Transposase [*Mycobacterium tusciae*] WP_040538977.1; 577aa 316/597(53%); 371/597(62%)
NZ_ANGC02000005.1	D092_RS15555	411922..412770	Site-sepcific integrase	Site-specific integrase [*Nocardia testacea*] WP_051165467.1; 410aa 165/312(53%); 207/312(66%)
NZ_ANGC02000007.1	D092_RS17655	107090..108331	Mobile element protein; integrase	Integrase [*Mycobacterium intracellulare*] WP_009954429.1; 417aa 288/413(70%); 324/413(78%)
	D092_RS17945	168554..169801	Mobile element protein: transposase: is*1164*	Transposase [*Mycobacterium sp.* NAZ190054] KWX61743.1; 412aa 339/404(84%); 367/404(90%)
	D092_RS17950	169852..170079	Dna resolvase/recombinase(pseudogen)	
	D092_RS17975	175058..177511	Hypothethical protein; putative DNA integrase/recombinase	Integrase [*Mycobacterium*] WP_046182709.1; 821aa484/792(61%); 574/792(72%)
	D092_RS17980	177508..178743	Putative DNA integrase/recombinase	Integrase [*Mycobacterium kumamotonense*] OBY29445.1; 395aa 277/390(71%); 313/390(80%)
	D092_RS18300	243517..244764	Mobile element protein: transposase: is*1164*	Transposase [*Mycobacterium sp.* NAZ190054] KWX61743.1; 412aa 339/404(84%); 367/404(90%)
NZ_ANGC02000011.1	D092_RS20485	7786..9111	Mobile element protein: transposase	Transposase [*Propionibacterium freudenreichii*] WP_048769214.1; 441aa 440/441(99%); 440/441(99%)
	D092_RS20525	16392..17669	Mobile element protein: transposase	Transposase [Jiangella muralis] WP_053208317.1; 416aa 387/421(92%); 399/421(94%)
	D092_RS20545	19950..20843	Integrase	Integrase [*Streptomyces zinciresistens*] WP_007505190; 296 aa 237/297(80%); 255/297(85%)

	D092_RS20550	20840..21169	Mobile element protein: transposase	Transposase [*Streptomyces sp.* NBRC 110035]; WP_042172217.1; 109 aa 82/109(75%); 91/109(83%)
	D092_RS20565	22392..23345	Putative integrase/recombinase	Recombinase [*Nocardia farcinica*] WP_060594881.1; 317aa 316/317(99%); 316/317(99%)
	D092_RS20610	29738..32749	Dde transposase	DDE transposase [*Gordonia terrae*] WP_004018622.1; 1003aa 974/1003(97%); 987/1003(98%)
	D092_RS20840	90718..91287	Phage protein	Hypothetical protein HMPREF0305_12394 [Corynebacterium pseudogenitalium ATCC 33035] EFQ79380.1; 137 aa 69/137(50%); 87/137(63%)
NZ_ANGC02000011.1	D092_RS20850	93198..93815	Resolvase/integrase bin	Putative resolvase [*Nocardia farcinica* IFM 10152] BAD60721.1; 222aa 197/205(96%); 200/205(97%)
	D092_RS20875	97358..100372	Dde transposase	DDE transposase [*Nocardia farcinica*] WP_011212398.1; 1005aa 971/1005(97%); 984/1005(97%)
	D092_RS20895	103583..105511	Phage integrase	Phage integrase (modular protein) [*Tetrasphaera australiensis* Ben110] CCH73768.1; 842aa. 326/641(51%); 427/641(66%)
	D092_RS21010	127468..127950	Transposase, pseudogen	Putative integrase [*Gordonia sihwensis* NBRC 108236] GAC62749.1; 398aa 257/390(66%); 302/390(77%)
NZ_ANGC02000014.1	D092_RS22495	78212..79378	Integrase	Transposase [*Nocardia elegans*] WP_063031757.1; 333aa282/332(85%); 301/332(90%)
	D092_RS22515	83731..84732	Mobile element protein: transposase	Transposase [*Corynebacterium maris*] WP_020935232.1; 384aa 172/262(66%); 195/262(74%)
NZ_ANGC02000020.1	D092_RS24085	35960..>36759	Mobile element protein: transposase	Transposase [*Corynebacterium maris]* WP_020935232.1; 384aa 172/262(66%); 195/262(74%)
NZ_ANGC02000021.1	D092_RS24110	5698..7239	Hypothetical protein; mobile element protein	Integrase [*Mycobacterium heraklionense*] WP_047318348.1; 513aa 379/513(74%); 411/513(80%)
	D092_RS24115	7306..8070	Atp-binding protein; mobile element protein	Transposase [*Mycobacterium canettii*] WP_015294405.1; 251aa 213/252(85%); 233/252(92%)
	D092_RS24135	12110..13276	Integrase	Putative integrase [*Gordonia sihwensis* NBRC 108236] GAC62749.1; 398aa 259/386(67%); 300/386(77%)
NZ_ANGC02000028.1	D092_RS24680	342..1487	Putative integrase/recombinase	Integrase [*Arthrobacter sp.* H14] WP_026536562.1; 365aa 229/361(63%); 274/361(75%)

Table S3. List of recombinases identified in the *R. ruber* Chol-4 genome.

Conting	Locus Tag	Localization	Protein Reference /size	Function	Max identity; Size; % identity; % Similarity
NZ_ANGC02000001.1	D092_RS02770	581107..582270	KXF88421.1 387AA	Rmuc family DNA recombination protein. The function of the rmuc protein is unknown but it is suspected that it is either a structural protein that protects	DNA recombinase [*Nocardia terpenica*] KZM71360.1; 378aa 266/382(70%); 300/382(78%)
	D092_RS02875	607260..609128	KXF88019.1 622AA	DNA against nuclease action, or is itself involved in DNA cleavage at the regions of DNA secondary structures	
	D092_RS02865	600680..603973	KXF88017 1097AA	Exodeoxyribonuclease v subunit alpha;recombinase recd; exodeoxyribonuclease v subunit gamma: Double-strand break repair via homologous recombination ; recbcd-dependent single-strand annealing (SSA).	Multispecies: exodeoxyribonuclease v subunit alpha [*Rhodococcus*] WP_054248544.1: 622aa 477/619(77%), 529/619(85%).
NZ_ANGC02000002.1	D092_RS08360	793035..794066	KXF87439.1 643AA	DNA recombinase	Multispecies: DNA recombinase [*Rhodococcus*] WP_017682306.1; 343aa 343/343(100%); 343/343(100%)
	D092_RS08085	731086..732003	KXF87397.1 305AA	Recombinase xerd; site-specific tyrosine recombinase xercd:it functions in circular chromosome separation.	Site-specific tyrosine recombinase xerd [*Mycobacterium* sp. Ga-2829]; WP_059089360.1: 317aa 232/305(76%); 258/305(84%)
	D092_RS07920	693426..693779	KXF87369 117AA	Hp;reca/rada recombinase: homologous recombination and underpin genome stability, by promoting the repair of double-stranded DNA breaks and the rescue of collapsed DNA replication fork.	Hypothetical protein [*Nocardia paucivorans*] WP_040790284.1; 113aa 98/117(84%); 105/117(89%)
NZ_ANGC02000003.1	D092_RS11460	670961..671842	KXF86722.1 293AA	Recombinase recb;exonuclease recb: homologous recombination.	Recombinase b [*Mycobacterium abscessus*] cpv10095.1; 293aa 178/282(63%) 210/282(74%)
	D092_RS09725	298516..299094	KXF86416 192AA	Recombination regulator recx; recombinase recx: recx inhibits reca recombinase and coprotease.	Recombination regulator recx [*Nocardia otitidiscaviarum*] WP_029928300.1; 166aa111/168(66%); 129/168(76%)
	D092_RS09720	296375..298507	KXF86415.1 710 AA	Recombinase reca, intein-containing: repair and maintenance of DNA	Recombinase reca, intein-containing [*Nocardia sp.* Bmg51109] WP_036567761.1; 711aa 628/711(88%); 667/711(93%)
	D092_RS09250	192362..193288	KXF86336.1 308AA	Recombinase xerc: site-specific tyrosine recombinase xercd.	Recombinase xerc [*Mycobacterium* sp. 852013-51886_sch5428379] obb61433.1; 300aa 205/296(69%); 237/296(80%)
NZ_ANGC02000005.1	D092_RS15475	392877..394130	KXF85732.1 417AA	Dna replication/repair protein recf;recombinase recf; (recfr, involved in recombination processes in mycobacteria)	Dna replication/repair protein recf *[Nocardia pseudovaccinii]* WP_063042415.1; 395aa 268/393(68%); 312/393(79%)
NZ_ANGC02000007.1	D092_RS18120	206614..207222	KXF85034.1 202AA	Recombination protein recr;recombinase recr (recfr, involved in recombination processes in mycobacteria)	Recombination protein recr [*Nocardia* sp. Bmg111209] WP_019931781.1; 202aa 190/202(94%); 197/202(97%)
	D092_RS17765	131651..132259	KXF84976.1 202AA	Recombination protein recr;recombinase recr	[repetido]
NZ_ANGC02000013.1	D092_RS21935	61919..63511	KXF84177 530AA	Recombinase recb Recbcd-dependent single-strand annealing (ssa)	Recombinase recb [*Nocardia tenerifensis*] WP_051186107.1; 537aa 334/529(63%); 394/529(74%)

Table S4 List of restriction modification systems identified in the *R. ruber* Chol-4 genome.

Contig	Locus Tag	Localization	Protein reference and size	Function	Max ID source/aa id/pos id
NZ_ANGC02000001.1	D092_RS03120	656070..657014	KXF88067 314aa	Mrr restriction system protein (endonuclease IV)	Restriction endonuclease [*Mycobacterium* sp. Ga-1199] WP_064350384.1; 295aa 149/296(50%); 197/296(66%)
	D092_RS03645	771133..771903	KXF88439 256aa	Endonuclease IV	Deoxyribonuclease IV [*Smaragdicoccus niigatensis*] WP_018159685.1; 257aa 183/253(72%); 212/253(83%)
	D092_RS04610	994228..994896	KXF88458 222aa	Deoxyribonuclease	Hypothetical protein duf1524 [*Nocardia nova* sh22a] ahh16003.1; 218aa 134/207(65%); 158/207(76%)
NZ_ANGC02000003.1	D092_RS08630	58570..59193	KXF86763 312aa	HNH endonuclease	Multispecies: HNH endonuclease [*Mycobacterium*] WP_023986061.1; 214aa 135/199(68%); 160/199(80%)
	D092_RS11355	641913..646058	KXF86705 1381aa	Endonuclease	Endonuclease [*Nocardia pseudovaccinii*] WP_063037236.1; 1403aa 855/1385(62%); 1015/1385(73%)
	D092_RS09220	186939..187295	KXF86330 118aa	Endonuclease	Hypothetical protein [*Nocardia grenadensis*] WP_063041706.1; 118aa 71/118(60%); 85/118(72%)
NZ_ANGC02000007.1	D092_RS17260	26161..26955	KXF84887 264aa	Endonuclease III	Endonuclease III [*Gordonia amarae*] WP_040514706.1; 268aa 209/257(81%); 222/257(86%)
	D092_RS17490	71233..71598	KXF84930 121aa	HNH endonuclease	Hnh endonuclease [*Modestobacter caceresii*] WP_036340911.1; 130aa 73/118(62%); 81/118(68%)
	D092_RS18325	248313..249338	KXF85067 341aa	Deoxyribonuclease	Deoxyribonuclease [*Gordonia* sp. Ucd-tk1] WP_065630560.1; 360aa 211/353(60%); 231/353(65%)
NZ_ANGC02000008.1	D092_RS18900	107380..108993	KXF84792 537aa	Type I restriction-modification system, DNA-methyltransferase subunit M (EC 2.1.1.72)	Type I restriction-modification system, DNA-methyltransferase subunit m [A*rthrobacter* sp. Pamc 25486] aiy01962.1; 539aa 424/538(79%); 466/538(86%)
NZ_ANGC02000010.1	D092_RS19995	23280..24380	KXF84462 366aa	HNH endonuclease	Hnh endonuclease [*Mycobacterium smegmatis*] WP_058125483.1; 359aa 211/352(60%); 254/352(72%)
NZ_ANGC02000011.1	D092_RS20810	78083..81217	KXF84373 1044aa	Type I restriction-modification system, restriction subunit R (EC 3.1.21.3)	Restriction endonuclease subunit R [*Rocardia arizonensis*] WP_054811818.1; 1044aa 965/1044(92%); 1009/1044(96%)
	D092_RS20820	82443..84188	KXF84375 581aa	restriction endonuclease subunit M (type I)	Restriction endonuclease subunit M [*Kutzneria* sp. 744] WP_043722299.1; 581aa 546/581(94%); 561/581(96%)
NZ_ANGC02000014.1	D092_RS22520	85260..87227	KXF84108 655aa	restriction endonuclease subunit S (Type I)	Restriction endonuclease subunit s [A*rthrobacter* sp. Yc-rl1] WP_047119061.1; 657aa505/647(78%); 570/647(88%)
NZ_ANGC02000018.1	D092_RS23505	1752..2426	KXF83805 224aa	Endonuclease	Endonuclease [*Gordonia araii*] WP_040521252.1; 227aa 198/227(87%); 208/227(91%)

CAPÍTULO 2

Functional differentiation of 3-ketosteroid Δ^1-dehydrogenase isozymes in *Rhodococcus ruber* strain Chol-4

1. Background

Steroids are a source of contamination of soil and waters and their presence has been detected even in drinking water, threatening many ways of life and public health (Soto, Calabro *et al.* 2004, Barel-Cohen, Shore *et al.* 2006, Gracia, Jones *et al.* 2008). Rhodococci can be useful in this biodegradation field due to their metabolic versatility and steroids degradation capability. On the other hand *Rhodococcus* spp. are potential biotechnological tools (van der Geize and Dijkhuizen 2004, Yam and Okamoto 2011) as they can provide with key enzymes essential for certain reactions that yield industrial needed intermediaries such as 4-androstene-3,17-dione (AD) and 1,4-androstadiene-3,17-dione (ADD) (Malaviya and Gomes 2008). But before exploiting all the advantages the different rhodococci offer, it is essential to know how these bacteria degrade steroids and which enzymes are involved in this process.

In the general scheme of steroid degradation, there are two key enzymes that initiate the opening of the steroid ring: the 3-ketosteroid-Δ^1-dehydrogenase [4-ene-3-oxosteroid: (acceptor)-1-ene-oxoreductase; EC 1.3.99.4)], also known as KstD and the 3-ketosteroid 9α-hydroxylase [Androsta-1,4-diene-3,17-dione; EC 1.14.13.142], also known as KshAB (Petrusma, van der Geize *et al.* 2014). KstD is a flavoenzyme involved in the Δ^1-dehydrogenation of the steroid molecule leading to the initiation of the breakdown of the steroid nucleus by introducing a double bond into the A-ring of 3-ketosteroids (Itagaki, Hatta *et al.* 1990, Florin, Kohler *et al.* 1996). This flavoprotein converts 4-ene-3-oxosteroids (e.g. AD) to 1,4-diene-3-oxosteroids (e.g. ADD) by trans-axial elimination of the C-1(α) and C-2(β) hydrogen atoms (Itagaki, Wakabayashi *et al.* 1990). KstD homologs have been identified in 100 different bacterial species (78 actinobacteria, 20 proteobacteria and 2 firmicutes) and at least in one fungus, *Aspergillus fumigatus* CICC 40167 (Chen, Wang *et al.* 2012, Kisiela, Skarka *et al.* 2012). Most of these KstD-containing bacteria occur in soil, marine or river sediments and are also able to degrade polycyclic aromatic hydrocarbons (Hilyard, Jones-Meehan *et al.* 2008). Phylogenetic analysis leads to classify the KstD-like enzymes in at least 4 different groups, in which KstD1, KstD2, KstD3 of *R. erythropolis* SQ1 are representatives of three of them (Knol, Bodewits *et al.* 2008). The crystal structure of the enzyme KstD1 of *R. erythropolis* SQ1 has been elucidated (Rohman, van Oosterwijk *et al.* 2013) confirming the presence of the two domains previously described, namely a N-terminal flavin adenine dinucleotide (FAD) binding motif and a substrate-binding domain (Wierenga, Terpstra *et al.* 1986, Molnár, Choi *et al.* 1995, Florin, Kohler *et al.* 1996, Knol, Bodewits *et al.* 2008).

The substrate range of different KstD proteins has been studied in *R. erythropolis* SQ1, being 3-ketosteroids with a saturated A-ring (e.g. 5α-androstane-3,17-dione and 5α-testosterone) the preferred substrates for KstD3 and (9α-hydroxy-)4-androstene-3,17-dione the favourite one for both KstD1 and KstD2 (Knol, Bodewits *et al.* 2008).

We have previously reported the occurrence of three KstD enzymes in *R. ruber* (NCBI: AFH57399 for KstD1; NCBI: AFH57395 for KstD2 and NCBI: ACS73883 for KstD3) (Fernández de las Heras, van der Geize *et al.* 2012). Growth experiments with single, double or triple *kstD* mutants proved that KstD2 is a key enzyme in the transformation of both AD to ADD and 9α-hydroxy-4-androstene-3,17-dione (9OH-AD) to 9α-hydroxy-1,4-androstadiene-3,17-dione (9OH-ADD) while both KstD2 and KstD3 are involved in the cholesterol catabolism in *R. ruber*. On the other hand, the role of KstD1 on the steroids catabolism remains unclear as *kstD1* mutation did not affect growing of this strain in steroids (Fernández de las Heras, van der Geize *et al.* 2012). In this study, we cloned the three *kstD* ORFs and heterologously expressed them in *R. erythropolis* CECT3014, in order to initiate the biochemical characterization of the encoded enzymes, as the basis for further studies on their applications. The results revealed that KstD3 uses more actively substrates with a saturated ring in contrast to KstD1 and KstD2. Additionally, we located and functionally defined the promoters of the three *kstD* ORFS in order to provide a basis for future research on the regulation of these genes.

2. Results and Discussion

As we have published earlier, the *R. ruber* strain Chol-4 genome contains three putative 3-ketosteroid Δ^1-dehydrogenase ORFs (*kstD1*, *kstD2* and *kstD3*) that code for flavoenzymes involved in the steroid ring degradation (Fernández de las Heras, van der Geize *et al.* 2012). Growth experiments with *kstD* mutants proved that KstD2 is the main enzyme involved in the transformation of AD to ADD. *R. ruber kstD2* mutants accumulate 9OH-AD from AD due to the action of a 3-ketosteroid-9α-hydroxylase (KshAB). On the other hand, only the strains lacking both KstD2 and KstD3 were unable to grow in minimal medium with cholesterol as the only carbon source. In order to know more about these *R. ruber* enzymes, we have performed transcriptional studies and followed their heterologous expression in *R. erythropolis* to detect their activities on a set of different substrates.

2.1. *In silico* analyses of *kstD* promoter regions of *R. ruber* strain Chol-4

A scheme of the three *R. ruber kstD* ORFs and their genomic surroundings is depicted in Figure. 1 showing also the *in silico* predicted pP1, pP2, pP3, pP4 and pP5 promoters. The available

programs (see Materials and Methods, section 2.5) yielded no putative promoter region just upstream either *kstD1* or *kstD2* ORFs, although it should be noted here that promoter prediction programs are not specific for Gram-positive species.

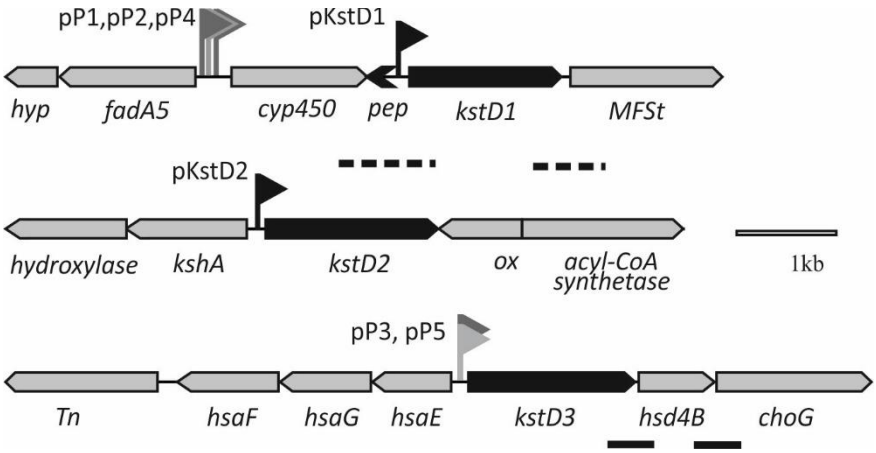


Figure 1. Schematic representation of *R. ruber* strain Chol-4 DNA *kstD* regions.
Putative promoters (pP) predicted using the BPROM (pP1: TTCCTT^{-35}...TGCTTGAAT^{-10}), PePPER (pP2: TTGAATGCTTTTAGAACGTGTTCCACATCgcgaC^{+1} and pP3: TGGACTCACCGCGCCATCATTCTATAACgtgtT^{+1}) or NNPP programs (pP4: GGTTGTCGTGGCGGACAAGGTGTGGTCCGAATGATCGGGAC^{+1}TTGGCGATT and pP5: GAAGGGATGGACTCACCGCGCCATCATTC^{+1}TATAACGTG) are shown in grey flags. TSS derived promoters (pKstD1, pKstD2) appear in black flags. Positive (solid line) and negative (dotted line) results of the amplification of co-transcribed products are depicted. Abbreviations**:** *hyp*: hypothetical protein; *fadA5*: acetyl-CoA acyltransferase; *cyp450*: cytochrome P450; *pep*: hypothetical peptide; *MFSt*: Major Facilitator Superfamily transporter; *kshA*: 3-ketosteroid 9α-hydroxylase; *ox*: oxidoreductase; *Tn: transposase*; *hsaF*: 4-hydroxy-2-oxovalerate aldolase; *hsaG*: acetaldehyde dehydrogenase; *hsaE*: 2-hydroxypenta-2,4-dienoate hydratase; *hsd4B*: 2-enoyl acyl-CoA hydratase; *choG*: cholesterol oxidase.

However, these programs detected putative promoters (pP1, pP2 and pP4) upstream the *cyp450* gene, lying in the intergenic region of the *fadA5-hyp* and *cyp450-kstD1-MFSt* opposite clusters. Flanking the pP1 putative promoter there are two palindromic sequences (TagAACagGTTgtc and TagAACgtGTTccA) (Figure 2), one of them rather similar and the second one identical to the consensus binding region reported for the KstR regulatory protein of *Mycobacterium* (TnnAACnnGTTnnA) (Kendall, Withers *et al.* 2007). KstR and KstR2, two TetR family repressor regulators, have been found to control most of the steroid pathways in actinobacteria (Kendall, Withers *et al.* 2007, Uhía, Galán *et al.* 2011, Crowe, Stogios *et al.* 2015). KstR is a highly conserved TetR family repressor that regulates the transcription of genes related to the upper and central pathway of cholesterol catabolism, namely the membrane transport of cholesterol, the degradation of the steroid side chain and the opening of the A and B rings (Kendall, Withers *et al.* 2007, Uhía, Galán *et al.* 2012). Upon binding to a 3-oxo-4-cholestenoic acid or to a CoA

thioester cholesterol metabolite, KstR releases the DNA and allows transcription to begin (García-Fernández, Medrano *et al.* 2014, Ho, Dawes *et al.* 2016). The KstR binding motif has been proved to be conserved within actinobacteria (Kendall, Withers *et al.* 2007, Shtratnikova, Schelkunov *et al.* 2016).

The possible co-transcription of the *cyp450-kstD1-MSFt* ORF cluster to a polycistronic mRNA would come into contradiction with the transcription of a non-yet described short putative ORF (*pep*, Figure 1) located in opposite sense in the 425 bp *cyp450-kstD1* intergenic region. This short ORF might code for a 35 amino acid peptide which shows a 98 % amino acid identity with part of the hypothetical 78 aa protein RHRU231_750039 (*R. ruber*, 78 aa). Therefore, *kstD1* might be independently transcribed while pP1/pP2/pP4 putative promoters might be only involved in *cyp450* transcription. Moreover, a putative ribosome binding site (GAAAGG) was found 9 bp upstream the *kstD1* initiation codon (Figure 2) that is identical to the proposed one for *sigA* of *R. ruber* TH (Ma and Yu 2012).

In the case of the *kstD2* ORF, none of the online programs recognized a promoter consensus, although this region contains some quasi-palindromic sequences and a Shine-Dalgarno-like motif (AGGAGC) (Figure 2).

There are two putative promoters for the *kstD3* ORF (pP3 and pP5, see Figure1 and 2) that lie in the intergenic region between *hsaE* and *kstD3* ORFs. The putative promoter pP3 contains the sequence TATAAC similar to the -10 consensus motif described for *Mycobacterium smegmatis* promoters ($T_{100\ \%}A_{93\ \%}T_{50\ \%}A_{57\ \%}A_{43\ \%}T_{71\ \%}$) (Bashyam, Kaushal *et al.* 1996), and a -35 region (**TGGAC**T) that resembles the *E. coli* promoter consensus motif TTGACA. In this region, there are also a tandem of two putative KstR binding sequences around this promoter (TgcAACctGTTtcc and TatAACgtGTTctA), one quite similar and the other identical to the KstR binding consensus, in a similar way to the *cyp450* pP1/pP2 putative promoter (Figure 2). The arrangement of these promoter and regulatory sequences, lying between opposite cluster genes, also occurs in the *R. ruber kstD1* region and it is similar to that found in *Mycobacterium* and *R. jostii* RHA1 genomes (Kendall, Withers *et al.* 2007, Kendall, Burgess *et al.* 2010).

On the other hand, Shell *et al.* have recently described that the abundance of leaderless transcripts (that lack a 5' UTR and a Shine-Dalgarno sequence and that begin with ATG or GTG) is a major feature of mycobacterial that accounts for around one-quarter of the transcripts (Shell, Wang *et al.* 2015). *kstD3* could be a leaderless ORF: there is no evidence of a Shine-Dalgarno sequence in its 5' region and the putative promoter is quite near to the ATG initiation

codon. Moreover, as it will be stated later, the promoter of this intergenic region is functional in *Rhodococcus* but not in *E. coli,* a fact that has also been confirmed in the mycobacterial leaderless messenger translation (Shell, Wang *et al.* 2015).

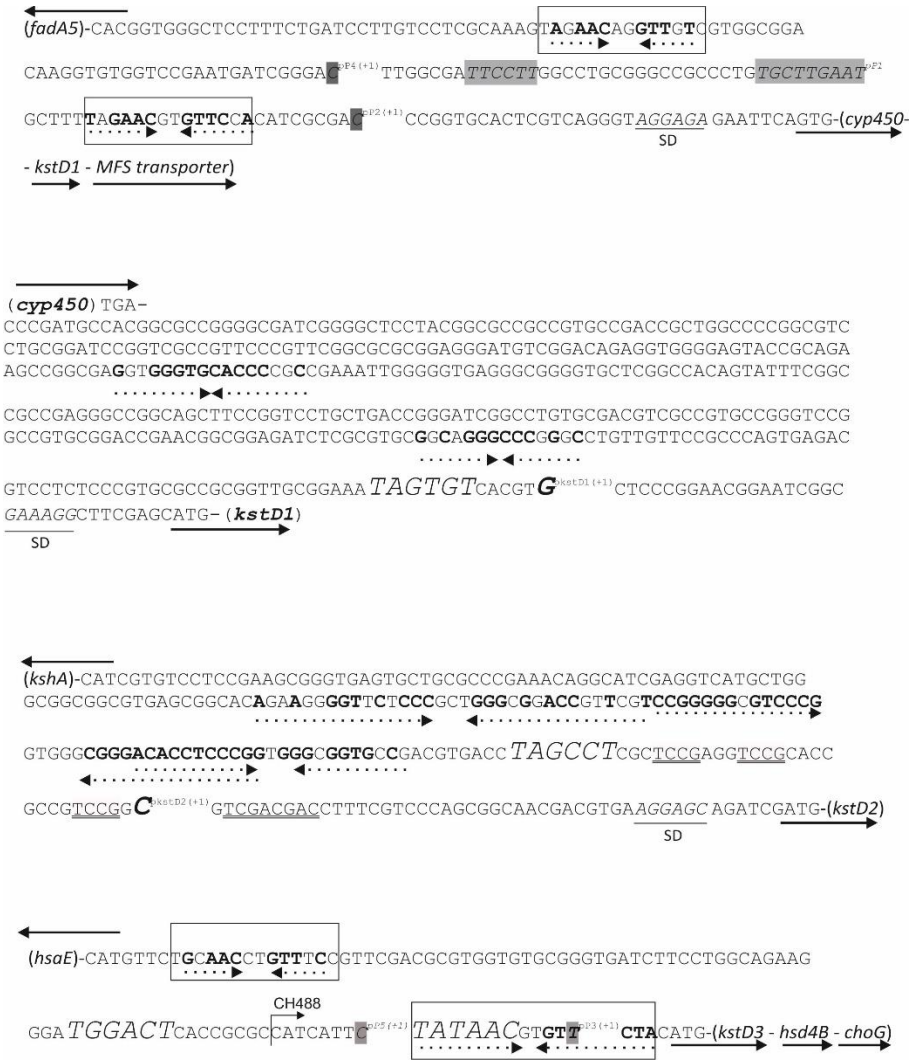


Figure 2. Sequence of regions upstream *kstD* ORFs. Solid arrows represent the orientation of the different ORFs from the initiation to the final codon. Sequences similar to *Mycobacterium* KstR binding sites (TnnAACnnGTTnnA) are within a square. Palindromic sequences appear in bold characters and dotted underlined. Shine Dalgarno (SD) sequences are in italics and underlined. The -10,-35 boxes are marked in grey. The *in silico* transcription initiation point of the putative pP1, pP2, pP3 and pP4 promoters are in italics and marked as pP(+1). The transcription start sites of *kstD1* and *kstD2* ORF obtained by the ARF-TSS method are shown in bold italics and marked as pKstD(+1). The TCCG repeats and the *Sal* box upstream *kstD2* are double underlined. Promoter signals similar to others described (e.g. *M. tuberculosis*) (Bashyam, Kaushal *et al.* 1996) appear in bigger size. Primer CH488 indicating the beginning of the minimum *kstD3* promoter is also shown.

2.2. Promoter cloning and characterization

To go further in the characterization of the promoter regions of the *R. ruber kstD* ORFs, a promoter-test vector suitable for *R. ruber* strain Chol-4 was constructed. *R. ruber* is sensitive to apramycin so we chose the expression of a gene encoding this resistance as a proof of the promoter activity.

pNV119 (Chiba, Hoshino *et al.* 2007), a *Nocardian* shuttle vector shown to replicate in *R. ruber* (Fernández de las Heras, van der Geize *et al.* 2012), was modified by adding the mcs of pSEVA351 that contains a transcriptional terminator in each extreme (Silva-Rocha, Martinez-Garcia *et al.* 2013, Durante-Rodriguez, de Lorenzo *et al.* 2014). The resulting pNVS plasmid (see Supplemental material 1) was used to study the activity of the Chol-4 putative promoter regions. The three *kstD* intergenic regions (Figure 2) plus the first 21 bases of each *kstD* ORF were PCR amplified, transcriptionally fused to the apramycin resistance gene obtained from pIJ773 and cloned into the mcs of pNVS. The recombinant plasmids were introduced into *R. ruber* by electroporation and kanamycin resistant clones were selected. *R. ruber* clones, harbouring the plasmids pNVSP1-A, pNVSP2-A or pNVSP3-A, were then plated in minimal medium supplemented with either 1.5 mM cholesterol, 1.5 mM AD or 10 mM sodium acetate, and in the presence of either 200 µg/mL kanamycin or 300 µg/mL apramycin. The apramycin resistance gene (Am^r) without any upstream promoter region was cloned in pNVS mcs generating the vector pNVSA that was used as a negative control. As a second control, a set of pNVSP vectors that contain every promoter region but do not carry the apramycin resistance gene was used. Figure 3 shows that only the cells harbouring the double system formed by a putative promoter and the apramycin resistance gene were able to grow on apramycin and kanamycin while cells harbouring the pNVSA or the pNVSPs vector were only able to grow in kanamycin plates. These results unambiguously confirm that all the three checked DNA regions contain *R. ruber* promoter sequences functionally active in the conditions used. On the other hand, *E. coli* harbouring the plasmids pNVSP1-A or pNVSP2-A were also able to grow in apramycin, in contrast with those harbouring pNVSP3-A (data not shown). Other actinobacteria promoters (e.g. some of *Mycobacterium* and *Streptomyces* spp.; (Strohl 1992, Bashyam, Kaushal *et al.* 1996) are also not functional in *E. coli* strains; this fact could related to the occurrence of leaderless genes (Shell, Wang *et al.* 2015).

Although none of the online programs recognized a promoter consensus for the *kstD2* ORF, the promoter-less vector pNVS has enabled us to check the promoter activity of the intergenic regions. The construction of an improved version of this plasmid that allows a quantitative

analysis of promoter strength is under work.

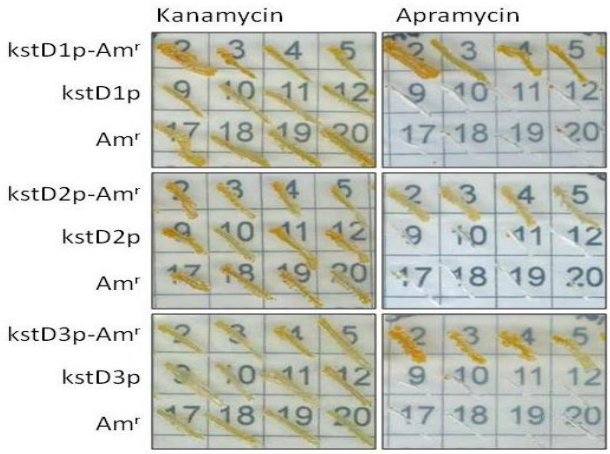


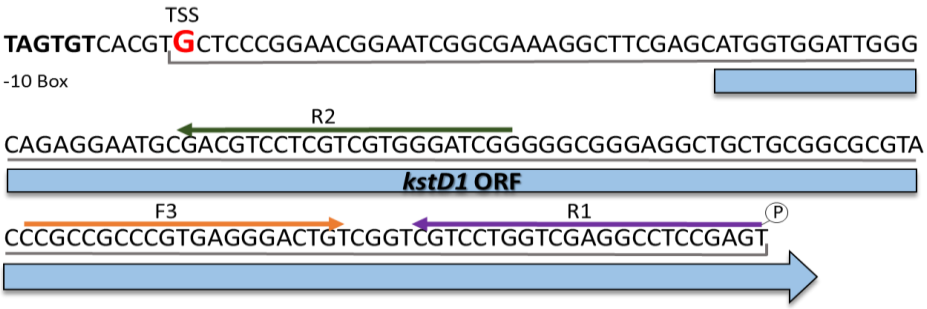
Figure 3. Comparative assessment of *R. ruber kstD* promoters. Cells of *R. ruber* strain Chol-4 harbouring different recombinant plasmids were grown in minimal medium supplemented with either 1.5 mM cholesterol (*kstD2, kstD3* or *kstD3*^{b} promoters) or 2 mM AD (*kstD1* promoter) in the presence of either 200 µg/mL kanamycin or 300 µg/mL apramycin. kstDp: *R. ruber* cells harbouring pNVSP plasmids (*kstD* promoters cloned in pNVS vector). kstDp-Am^{r}: *R. ruber* cells harbouring pNVSP-A plasmids (apramycin resistance gene fused to *kstD* promoters in pNVSP plasmids). Am^{r}: pNVS vector containing the promoter less apramycin resistance gene.

To define the transcription start sites (TSS) of the *kstD* genes, the transcription start point protocol (ARF-TSS) indicated in Materials and Methods, section 2.6 was followed in *R. ruber* cells. We could conclude that the 5' terminal base in *kstD1* messenger RNA is a G resulting from the transcription that starts 34 bp upstream the *kstD1* initiation codon. Similarly, the transcription start site of the *kstD2* gene is a C 48 bp upstream the *kstD2* initiation codon (Figure 2 and Figure 4). However, this approach did not yield any result in the case of the *kstD3* gene. In order to better define the limits of the *kstD3* promoter, progressively shorter sections of the intergenic region, keeping the first 21 bases of *kstD3* ORF, were PCR amplified and transcriptionally fused to the apramycin resistance gene.

As it can be seen in Figure 2 and Figure 3, just a minimum region of only 23 pb upstream the ATG of *kstD3* ORF (pNVSP3^{b}-A vector) was enough to act as a promoter as the *Rhodococcus* cells harbouring this vector were able to grow on medium with either apramycin or kanamycin. This promoter region is partially similar to the putative pP3 and pP5 promoters mentioned before, containing the TATAAC sequence similar to the -10 consensus motif described for *Mycobacterium smegmatis* promoters (Figure 1).

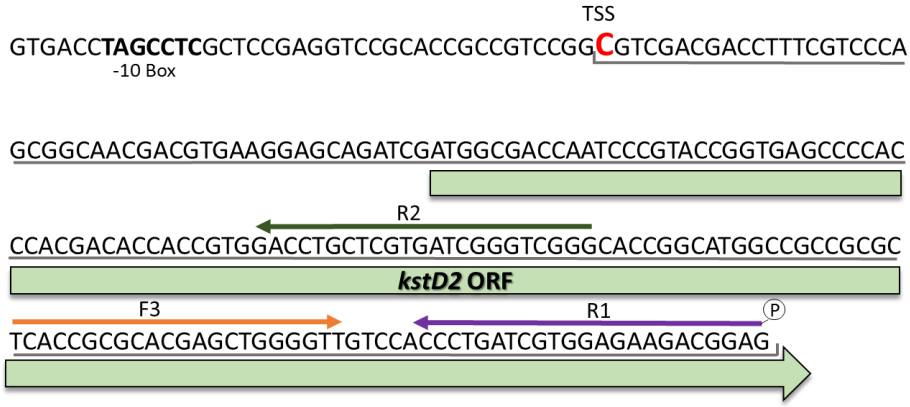
Putative −35 and −10 hexamers identification was based on the sequence of other actinobacteria promoters. The -10 region motif (TAGTGT) found 40 bases upstream the *kstD1*

initiation codon is similar to the -10 consensus region described for *Mycobacterium tuberculosis* promoters ($T_{80\ \%}A_{90\ \%}Y_{60\ \%}G_{40\ \%}A_{60\ \%}T_{100\ \%}$) and identical to the T101 promoter described by Bashyam *et al.* in the same bacteria (Bashyam, Kaushal *et al.* 1996). No sequence similar to any -35 motif was found in the *kstD1* region. The absence of -35 motifs seems to be a characteristic



of actinobacterial genes (Strohl 1992, Bashyam, Kaushal *et al.* 1996).

Figure 4. ARF-TSS analysis of *kstD1*and *kstD2* transcripts. The TSS identified by ARF-TSS appears in red and the proposed -10 box appear in black. R1, R2 and F3 primers are indicated. A grey bar shows the sequence obtained by ARF-TSS from the circularized cDNA and the subsequent PCR amplification with R2 and F3.



There is a putative -10 sequence (TAGCCT) in the intergenic region between *kshA* and *kstD2* that resembles the T6 promoter of *M. tuberculosis* (TAGGCT) (Bashyam, Kaushal *et al.* 1996). However, it is a bit far away from the *kstD2* TSS. There is a repetitive motif TCCG in this region (Figure 2) that also appears in the human *Sal* box element present in the 3' terminal spacer of rDNA and that constitutes a termination signal for RNA polymerase I (**TCCGCAC**GG**GTCGAC**C**A**G) (Pfleiderer, Smid *et al.* 1990). In this *kstD2* upstream region we found something similar: **TCCG**AGG**TCCGCAC**CGCCG**TCCG**GC**GTCGAC**G**A**C, where black-underlined letters marked the resemblance between them. Such promoter proximal terminators can appear also upstream of

the transcription start site and in this case, they are described to positively affect transcription initiation and to prevent transcriptional interference by reading through of polymerases from the spacer that separates each rDNA repetition (Grummt, Kuhn *et al.* 1986, Henderson and Sollner-Webb 1986, Henderson, Ryan *et al.* 1989). The biological significance of this motif in this intergenic *kstD2* region should be further determined.

2.3. Transcriptional analysis of *kstD* genes in *R. ruber* strain Chol-4

The transcription of the three *kstD* genes found in *R. ruber* strain Chol-4 was analysed by RT-PCR of RNA samples prepared from cultures grown in either M457 mineral medium or LB medium supplemented with either AD or cholesterol as possible inductors. Control cultures were grown in either M457 supplemented with sodium acetate (2 g/L NaAc) or LB, both in the absence of any steroid. Using the specific primer pairs designed to search for the transcript of each different ORF (see Materials and Methods, Table 4), we could show that all the *kstD* genes are transcribed in all the conditions used in our assays (Figure 5) and that transcription of some of them is induced by cholesterol or AD. There are also other *Rhodococcus* metabolic genes reported to show a low-level constitutive transcription that can be strongly induced under the presence of a determinate substrate. For instance, phenol degradation genes in *R. erythropolis* are constitutively transcribed and also highly induced by phenol (Szoköl, Rucká *et al.* 2014).

cDNA templates concentrations were adjusted to the same value in RT-PCR experiments, so we can consider the differences observed in the thickness of some amplification bands to be meaningful (Figure 5). Specific amplification of *kstD1* cDNA was higher in AD induced than in non-induced cultures, while amplification of *kstD3* cDNA was higher in cultures induced with cholesterol as compared to the other conditions assayed. So we can conclude that these two genes, although constitutively transcribed, they can also be additionally induced by the presence of AD (*kstD1*) and cholesterol (*kstD3*). The induction of *kstD* genes by cholesterol or AD has also been reported in other microorganisms: a 3.3 expression ratio (cholesterol/pyruvate) was reported for one *kstD* in *R. jostii* RHA1 (*ro04532*) (van der Geize, Yam *et al.* 2007, Mathieu, Mohn *et al.* 2010), while other putative *kstD* genes of the same strain (e.g. *ro02483*, *ro05798* and *ro05813*) were up-regulated in 7-ketocholesterol but not in cholesterol (Mathieu, Mohn *et al.* 2010); the main *kstD* in *M. smegmatis* (*MSMEG_5941*) was 13-fold up-regulated in cholesterol respect to glycerol (Uhía, Galán *et al.* 2012); a 1.8, 4.1 or 1.2 -fold up-regulation (AD/glycerol) was found for *kstD1*, *kstD3* and *kstD2* genes respectively in *Mycobacterium neoaurum* (Yao, Xu *et al.* 2014). These differences in regulation among the different *kstD* genes within the same strain highlight the view that KstD proteins may be acting in different metabolic steps and/or

pathways, each one having a particular catalytic role.

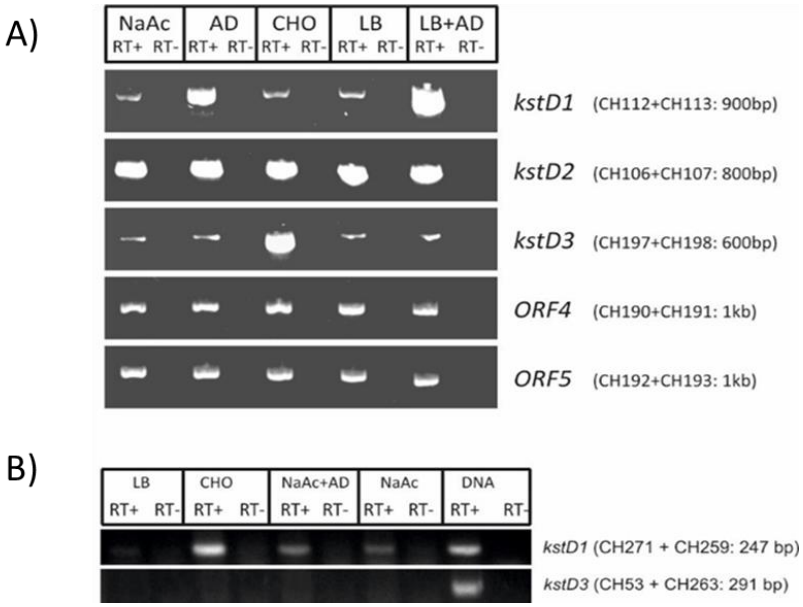


Figure 5. A). Analysis of transcription products of *kstD* genes in *R. ruber* strain Chol-4 by specific RT-PCR amplification and agarose gel electrophoresis (Published in Fernández de las Heras, 2013). RNA samples were isolated from cultures grown in: M457 minimal medium supplemented with sodium acetate (lane NaAc), AD (lane AD) or cholesterol (lane CHO) and LB medium supplemented with (lane LB+AD) or without AD (lane LB). RT– refers to negative controls (not incubated with retrotranscriptase) to exclude DNA contamination in RNA samples. Primers used in each PCR reaction and amplified fragment size are shown in brackets. As non-induced expression controls, the transcription of two *R. ruber* Chol-4 enzyme genes was followed in the same conditions: ORF4, that codes for a FAD-binding dehydrogenase (99 % identity with the protein NCIB: WP_010594120.1 of *Rhodococcus* sp. P14) andORF5, that codes for a fumarate reductase (99 % identity with NCIB: WP_010594021.1 of *Rhodococcus* sp. P14; see sequence in Supplemental material 2. **B)**. RT-PCR assays of *R. ruber* Chol-4 *kstD1* and *kstD3* transcription carried out in *kstD2 R. ruber* mutant growing in: LB (lane LB) or M457 plus CHO (lane CHO), NaAc and AD (lane NaAc+AD) or NaAc (lane NaAc). Lane DNA shows the result of the control experiment: PCR made on *R. ruber* DNA.

The amplification of *R. ruber* strain Chol-4 *kstD2* cDNA yielded thick amplification bands in all cases (Figure 5), which clearly leads to propose a constitutive expression of *kstD2*. In a similar way, two putative *kstD* genes in *R. jostii* RHA1 genome (*ro90203* and *ro09040*, belonging to the KstD2-branch of the KstD phylogenetic tree) (Knol, Bodewits *et al.* 2008) were expressed but not up-regulated neither in cholesterol nor in 7-ketocholesterol, when compared to pyruvate (Mathieu, Mohn *et al.* 2010).

The expression profile of R. ruber kstD1, kstD2 and kstD3 genes was determined by real-time PCR. Taking as 1 the expression levels on sodium acetate (unexposed steroid culture), the values obtained for the expression of the three genes were: 7.6, 2.0 and 240.5-fold for kstD1, kstD2 and kstD3, respectively, in cultures grown in cholesterol; and 13.6, 0.7 and 0.6-fold for kstD1,

kstD2 and kstD3, respectively, in cultures grown in AD.

The particular organization of *R. ruber* Chol-4 *kstD1* and *kstD3* genes (Figure 1) opens the possibility that polycistronic *kstD* mRNAs could be synthesized by the co-transcription of the *cyp450-kstD1-MFS transporter* and *kstD3-hsd4B-choG* gene clusters. The results of the RT-PCR experiments did not show the occurrence in AD culture medium of either *cyp450-kstD1* or *kstD1-MFS transporter* RNA sequences, indicating that *kstD1* gene is independently transcribed.

In contrast, co-transcripts from both *kstD3-hsd4B* ORFs and *hsd4B-choG* ORFs could be amplified from cultures grown in the presence of cholesterol, strongly suggesting that *kstD3-hsd4B-choG* ORFs are co-transcribed into a polycistronic mRNA (Figure 1). This group of three genes are also described to be co-transcribed in *R. erythropolis* (Fernández de las Heras, Mascaraque *et al.* 2011).

The adjacent location of *kstD3-hsd4B* ORFs is highly conserved among rhodococci. The *hsd4B* ORF encodes a 2-enoyl acyl-CoA hydratase involved in the β-oxidative cycle of the C-17 cholesterol side chain (van der Geize, Yam *et al.* 2007). The *choG* ORF encodes an extracellular cholesterol oxidase that it is involved in the first step of cholesterol catabolism that implies its conversion to 4-cholesten-3-one (Kreit and Sampson 2009, Pollegioni, Piubelli *et al.* 2009, Vrielink and Ghisla 2009, Fernández de las Heras, Mascaraque *et al.* 2011, Fernández de las Heras, Perera *et al.* 2014). All this suggests that *kstD3, hsd4B* and *choG* genes are mainly involved in the steroid catabolism.

In this work we showed that *kstD2* is constitutively and strongly transcribed, while *kstD1* and *kstD3* are also constitutively but faintly transcribed, although they can be highly induced by AD or CHO, respectively. In a previous work (Fernández de las Heras, van der Geize *et al.* 2012), we reported the construction of a *kstD2* deletion mutant of *R. ruber* that is unable to grow in minimal medium supplemented with AD. The question is why KstD1 and/or KstD3 cannot substitute KstD2 allowing the *kstD2* deletion mutant to grow on AD. To partially advance in this subject, RT-PCR experiments were also performed on RNA from the *kstD2 R. ruber* mutant to check the expression of the other *kstD* genes (Figure 5B).

The results showed that the transcription pattern of the *kstD* genes of the mutant is far different than that of the same genes in the wild type. Namely, *kstD1* gene in the mutant was constitutively and slightly transcribed and its expression is induced in CHO. A more noticeable change affects to the *kstD3* gene, which is not transcribed at all in any of the conditions used. These data reveal a complex relationship among the KstD enzymes and their expression control

mechanisms.

Modification of the *kstD* transcription levels of genes remaining in the cell have also been described in a *Mycobacterium neoaurum kstD* mutant: the transcription ratio of the *kstD1* ORF (similar to the *kstD3* ORF from *R. ruber*) in AD induced cultures respect to glycerol cultures increases from 1.8 (in the wild type strain) to 2.7-fold (in the *kstD3 M. neoaurum* mutant strain) (Yao, Xu *et al.* 2014).

An appealing conclusion of that complex situation is that the three KstD proteins of *R. ruber* strain Chol-4 may be differentially involved in distinct pathways of steroid degradation, and that their expression could also be differentially and specifically controlled.

2.4. Heterologous expression of KstD1, KstD2 and KstD3 of *R. ruber*

The three *R. ruber kstD* genes (*kstD1*, *kstD2* and *kstD3*) were cloned into the pTip-QC1 expression vector. *R. erythropolis* CECT3014 cells were electroporated with these constructions and clones harbouring each of those recombinant plasmids were isolated. Expression of the KstD proteins from these vectors in the CECT3014 transformed cells was followed by SDS-PAGE analysis. Molecular weights were 54.8, 60.8 and 61.8 kDa for KstD1, KstD2 and KstD3 respectively (Supplemental material 3). Cell-free extracts of cultures grown from these clones were used for the analysis of KstDs activities, and the kinetic parameters of the heterologously expressed KstDs from *R. ruber* were followed for different substrates (Table 1). Control cell extract from the *R. erythropolis* culture harbouring an empty pTip-QC1 vector yielded none or very low basal levels when acting on all the substrates used in the assay and were taken into account for final activities.

The substrate profile of KstD1 showed a clear preference to 9OH-AD and testosterone, followed by progesterone, Deoxy corticosterone (DOC) and AD (Table 1). All these compounds display a keto group at C3, a C4-C5 double bond and an electronegative side-chain at C17. When comparing to the substrate preference order of *R. erythropolis* SQ1 KstD1 (Prog > 9OH-AD, AD > 5α-Tes> BNC > 11βCort), some details highlights: i) $KstD1_{SQ1}$ has a relative catalytic efficiency (RCE) on progesterone 3.4 times higher than on AD, a very big difference to the ratio 9OH-AD/AD (1.6) showed by *R. ruber* KstD1 (Table 1); ii) *R. ruber* KstD1 is not active on 5α-Tes in contrast to $KstD1_{SQ1}$.

The order of substrate preference of *R. ruber* KstD2 placed progesterone in the first position, followed by 5α-Tes, DOC, AD and testosterone (Table 1), displaying a substrate profile similar to

KstD2$_{SQ}$1 enzyme (Prog > AD > 5α-Tes > 9OH-AD, BNC > 11βCort). It is noteworthy the similarity of both KstD profiles, having in mind that they have been expressed in different cellular context (a *Rhodococcus* strain, and *E. coli*). *R. ruber* KstD2 has a broader range of substrates than *R. ruber* KstD1 as it can act on all the KstD1 substrates and also on 5α-Tes that contains a saturated A ring (Table 1).

R. ruber KstD3 did not show any activity at all when acting on AD or 9OH-AD, which are considered the natural substrates for KstD enzymes, but it showed the highest activity when using 5α-Tes as substrate followed by progesterone and lastly DOC (Table 1). *R. ruber* KstD3 has a very narrow substrate range similarly to the *R. erythropolis* SQ1 and *M. tuberculosis* H37Rv isoforms, being the A-ring saturated 5α-Tes the preferred substrate for all KstD3 enzymes (Table 1, (Knol, Bodewits *et al.* 2008)). However, KstD3 affinity for 5α-Tes differs from 33-36 μM in the cases of KstD3$_{H37RV}$ and KstD3$_{SQ1}$ respectively, to 181 μM in *R. ruber* KstD3 (Table 1). Despite this low affinity for 5α-Tes, this is the best substrate for the *R. ruber* enzyme among those that were assayed (Table 1).

On the other hand, although it has been proposed that only steroids carrying a small or no aliphatic side chain at C-17 are suitable substrates for KstD3 (Knol, Bodewits *et al.* 2008), a minor activity, not very different to that obtained with progesterone, has been observed in the assays of *R. ruber* KstD3 enzyme on DOC (Table 1). *R. ruber* KstD3 seems to be more related to the cholesterol metabolism than to the AD metabolism (Fernández de las Heras, van der Geize *et al.* 2012) and then it could be acting on some not yet neatly defined intermediaries of the bacterial cholesterol metabolism.

None of the three *R. ruber* KstD proteins displayed detectable activity on 4-cholestene-3-one, 5α-cholestane-3-one or 5β-androstane-3,17-dione (5β-AD), ADD , cholesterol, cholestenone, cholic acid, DHEA, ergosterol, stigmasterol, β-estradiol, sodium deoxycholate or 5-pregen-3-β-nolone. Therefore, these KstDs catalyse preferentially 4-ene-3-oxosteroids

TABLE 1: Substrate profiles of *R. ruber* KstDs expressed an analyzed in cell-free extracts of *R. erythropolis*.

	KstD1				KstD2				KstD3			
Substrate	Rel. Act %	Km (μM)	RCE	RCE/RCE_{AD}	Rel. Act %	Km (μM)	RCE	RCE/RCE_{AD}	Rel. Act %	Km (μM)	RCE	RCE/RCE_{AD}
AD	100.0±11.6	34.2±3.8	2.92	1.00	100.0±11.8	40.1±8.7	2.60	1.00	nd	nd		
9OH-AD	107.4±13.4	22.1±7.0	4.84	1.66	29.8±6.5	543.2±77.8	0.05	0.02	nd	nd		
4BNC	52.3±7.4	76.1±15.7	0.69	0.24	30.8±7.2	38.2±6.6	0.81	0.31	nd	nd		
Prog	89.7±7.9	27.9±9.5	3.22	1.10	182.0±7.0	33.8±4.9	5.38	2.07	18.6±3.7	43.6±2.9	0.42	0.78
Cort	20.3±5.8	161.6±7.6	0.13	0.04	19.0±3.0	374.3±74.5	0.05	0.02	nd	nd		
Tes	134.6±22.6	28.8±7.4	4.67	1.60	233.3±23.3	107.9±18.5	2.16	0.83	nd	nd		
19OH-AD	39.0±10.6	368.8±102.4	0.11	0.04	24.6±7.8	347.4±41.7	0.07	0.03	nd	nd		
DOC	67.0±7.4	21.6±4.5	3.10	1.06	124.2±17.2	42.5±5.8	2.92	1.12	21.6±8.8	111.3±4.0	0.19	0.36
5α-Tes	nd	nd			75.2±4.8	24.57±7.8	3.06	1.18	100.0±21.9	181.1±42.6	0.55	1.00

Rel. Act: relative activity values. Enzyme activities are expressed as percentage of activity of AD (for KstD1 and KstD2 with 3.2 U/mg and 1.4 U/mg respectively) or 5α-tes (for KstD3 with 0.3 U/mg) that were set as 100%. **nd**: enzyme activity was not detected for this substrate. **RCE**: relative catalytic efficiency given by the ratio Rel. Act/Km. **Prog:** progesterone; **Cort**: corticosterone; **Tes**: testosterone; **5α-tes:** 5-alpha-testosterone; **19OH-AD**: 19-Hydroxy-4-androstene-3,17-dione; **DOC**: Deoxycorticosterone, **4-BNC**: 4-pregnen-3-one-20β-carboxylic acid.

2.5. Molecular Modelling of KstD1 of *R. ruber* strain Chol-4

The sequence of the three KstDs of *R. ruber* were described previously (Fernández de las Heras, van der Geize *et al.* 2012). In an attempt to go deeper inside the reasons of the catalytic and kinetic differences among some of these enzymes, we performed protein sequence analysis and modelling studies on the *R. ruber* KstDs using different approaches. The published three-dimensional structure and the catalytic mechanism proposed for KstD1 from *R. erythropolis* SQ1 has been used as a suitable model (Petrusma, Hessels *et al.* 2011, Rohman, van Oosterwijk *et al.* 2013). The I-Tasser and PredictProtein programmes predicted that none of the three *R. ruber* KstDs contain sulphur bridges or transmembrane segments.

The catalytic mechanism of *R. erythropolis* SQ1 KstD1 is based in the keto-enol tautomerization of the substrate caused by Tyr^{487} and Gly^{491} residues that increases the acidity of the C2 hydrogen atoms of the substrate. Then Tyr^{119} and Tyr^{318} capture the axial β-hydrogen from C2 as a proton whereas the FAD molecule accepts the axial α-hydrogen from the C1 atom of the substrate as a hydride ion (Petrusma, Hessels *et al.* 2011, Rohman, van Oosterwijk *et al.* 2013). Table 2 collects the residues involved in the active site of $KstD1_{SQ1}$ described for Rohman *et al.* (Rohman, van Oosterwijk *et al.* 2013) and the homologue residues of the three KstDs from *R. ruber* found by COBALT programme.

Table 2: Homologous residues of KstDs implicated in the catalytic pocket attending to Rohman *et al.* (Rohman, van Oosterwijk *et al.* 2013). Grey files contain the four key residues of the KstD1 active site reported for *R. erythropolis* SQ1: Tyr^{119}, Tyr^{318}, Tyr^{487} and Gly^{491}.

KstD1-SQ1	KstD1-ruber	KstD2-ruber	KstD3-ruber
S^{52}	G^{53}	A^{59}	G^{50}
F^{116}	W^{118}	Y^{125}	Y^{116}
Y^{119}	Y^{121}	Y^{128}	Y^{119}
F^{294}	F^{295}	P^{337}	F^{327}
V^{296}	L^{297}	V^{339}	S^{330}
Y^{318}	Y^{319}	Y^{366}	Y^{354}
I^{352}	T^{356}	--	I^{395}
I^{354}	V^{358}	F^{405}	P^{397}
L^{447}	L^{447}	L^{499}	L^{490}
Y^{487}	Y^{487}	Y^{539}	Y^{530}
P^{490}	G^{490}	A^{542}	P^{533}
G^{491}	G^{491}	G^{543}	G^{534}
V^{492}	N^{492}	A^{544}	A^{535}
P^{493}	P^{493}	T^{545}	T^{537}

The four key residues: Tyr^{119}, Tyr^{318}, Tyr^{487} and Gly^{491} of the $KstD1_{SQ1}$ active site have a counterpart residue in the *R. ruber* KstDs. However, the I-TASSER model prediction of these *R. ruber* enzymes shows that the orientations of the side chain of the tyrosine residues differ within the catalytic pocket site being specific of each enzyme. The variation in orientation of the key residues inside the catalytic poked is shown in figure 6. Particularly, the orientation of KstD3 Y^{354} and Y^{530} residues is almost opposite to that of its homologous KstD1 and KstD2 residues. Moreover, amino acid Y128 from KstD2 has a position highly separated compared to its homologous from KstD1 (Y^{121}) and KstD3 (Y^{119}). These differences could justify the different affinity and catalytic properties among the three KstDs. The secondary structure and solvent accessibility prediction by PredictProtein of the three KstDs from *R. ruber* are gather in table 3. The current shortage of crystalline structures of these proteins greatly limit more detailed conclusions.

Table 3. PredictProtein prediction of the secondary structure and solvent accessibility in % of the three KstDs from *R. ruber*

Structure	**KstD1**	**KstD2**	**KstD3**
Strand	12.72	11.88	64.74
Loop	63.99	64.01	14.38
Helix	23.29	24.11	20.88
Accessibility	**KstD1**	**KstD2**	**KstD3**
Exposed	24.1	21.81	22.11
Buried	67.7	71.63	70.5
Intermediate	8.22	6.56	7.54

A redundancy of KstD and Ksh enzymes have been described in the actinobacteria genomes (Knol, Bodewits *et al.* 2008, Mathieu, Mohn *et al.* 2010, Petrusma, Hessels *et al.* 2011, Fernández de las Heras, van der Geize *et al.* 2012). This redundancy could provide the cell with a bigger metabolic versatility and a fine-tuned response to the challenging environment. In the case of KstD enzymes, three homologues have been found in *R. erythropolis* strain SQ1 and in *Mycobacterium neoaurum* that displayed different substrate preferences and that could be involved in different metabolic steps in a strain-dependent way (van der Geize, Hessels *et al.* 2002, Knol, Bodewits *et al.* 2008, Yao, Xu *et al.* 2014).

Even more, it has been shown recently that mutations in the KstDs provokes a different ADD/AD molar ratio (Shao, Zhang *et al.* 2016) and that environmental factors such as an increase of temperature can inhibit the KstD/Ksh action (Xu, Gao *et al.* 2015) on phytosterol in *Mycobacterium* sp. These multiplicity and versatility give to these enzymes a substantial role in

the catalytic dehydrogenation of several related steroid molecules and pose many difficulties to clarify the particular role and way of acting of every single KstD.

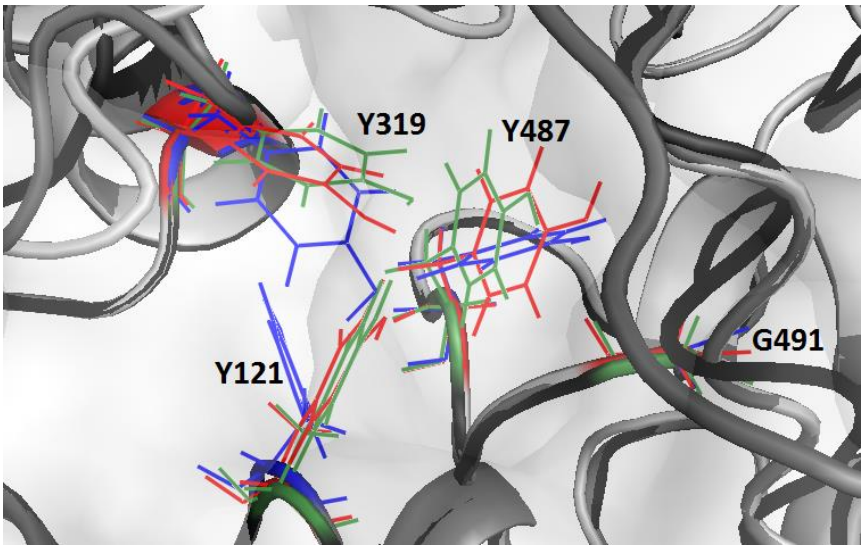


Figure 6. Modelling of the active site of KstDs. The orientation of the four key residues in the KstD binding pocket is shown. They are superposed and depicted in different colours: green for KstD1, blue for KstD2 and red for KstD3. Only KsD1 residues are named, their counterpart homologues residues of KstD2 and KstD3 being listed in Table 2.

Our results suggest that both KstD1 and KstD2 of *R. ruber* could act in the conversion of AD to ADD being KstD1 mainly involved in the 9OH-AD to 9OH-ADD conversion, in a similar way to what has been described in *R. erythropolis* SQ1 (van der Geize, Hessels *et al.* 2002). However, there are differences between these two strains, as the necessity of a double *kstD1* and *kstD2* mutation to prevent the growth in AD in the case of *R. erythropolis* SQ1 (van der Geize, Hessels *et al.* 2002) or *R. rhodochrous* DSM43269 (Liu, Shen *et al.* 2016), while the same effect is obtained by the single *kstD2* deletion in *R. ruber* strain Chol-4 (Fernández de las Heras, van der Geize *et al.* 2012), suggesting that this last mutation affects in some way the activity of the KstD1 protein.

kstD3 ORF occurs in the *R. ruber* genome in a quite conserved location within *Rhodococcus* species, clustered with *hsd4B* (which encodes a 2-enoyl acyl-CoA hydratase proposed to be involved in cholesterol side-chain shortening) (van der Geize, Yam *et al.* 2007) and *choG* (coding for a cholesterol oxidase that converts cholesterol into cholestenone) (Pollegioni, Piubelli *et al.* 2009, Fernández de las Heras, Mascaraque *et al.* 2011, Fernández de las Heras, Perera *et al.* 2014). Recently it has been proposed that sterols can be catabolised in *R. equi* USA-18 by two partially different pathways, namely via AD or via Δ1,4-BNC, that converge in the intermediary

9OH-ADD (Yeh, Kuo *et al.* 2014). Given its substrate preference, the fact that the growth in AD is independent of the KstD3 activity while the growth in cholesterol needs the presence of either KstD2 or KstD3 in *R. ruber* (Fernández de las Heras, van der Geize *et al.* 2012), lead us to suggest that KstD3 may be involved in an alternative AD-independent cholesterol catabolic pathway.

1.1. Growth of *R. ruber* strain Chol-4 ΔkstDs strains in different media

The ability of the *R. ruber* mutants to grow on different steroid media was checked by cultivating them in 457 DSMZ minimal medium supplemented with the desired carbon source (Table 4). Steroids were prepared in 16 mM cyclodextrins as indicated in methods (*R. ruber* was not able to grow in cyclodextrins as single carbon source). Table 4 shows the data obtained after a 48 h growth of the wild type and the mutant strain cultures.

Table 4: Growth of *R. ruber* wild type (WT) and *kstD* mutants on different substrates after 48 hours of culture

Carbon source/ Strain	CHO 1.6 mM	AD 2.2 mM	ADD 2.2 mM	Cholic 2.2 mM	Tes 2.2 mM	DHEA 2.2 mM	Phy 0.6 mg/ml	Acet 2 mM
ΔkstD1,2	++	-	++	-	+	+	++	++
ΔkstD1,3	++	++	++	++	++	+	++	++
ΔkstD2,3	+	-	++	-	+	+	+	++
ΔkstD1,2,3	+	-	++	-	+	+	+	++
WT	++	++	++	++	++	++	++	++

Cholic: cholic acid; **Tes**: testosterone; **DHEA**: dehydroepiandrosterone; **Phy**: phytosterols; **Acet**: sodium acetate. +: A_{600nm} < 0.3; ++: A_{600nm} >0.3.

These results show that KstD2 is the only enzyme involved in the AD catabolism as it was previously shown by Fernández de las Heras *et al.*, (2012). KstD2 also participates in the degradation of other substrates such as cholic acid, testosterone, cholesterol and phytosterols. In the case of long chain steroids, the isoform KstD3 is involved.

3. Conclusions

To sum up, this study provides biochemical and genetic insights into the three KstD proteins found in *R. ruber*. The kinetic differences between the three KstDs suggest that each enzyme could act on different steps of the steroid catabolic routes. Attending to the *in vitro* data, both KstD1 and KstD2 could be involved in the AD catabolism while KstD1 would have a preference for 9OH-AD and KstD2 for progesterone. KstD2 seems to be a more versatile enzyme than KstD1 in *R. ruber* as it can act also on saturated steroid substrates such as 5α-Tes. On the other hand, the narrower range of substrates for KstD3 and its preference for saturated steroid made this

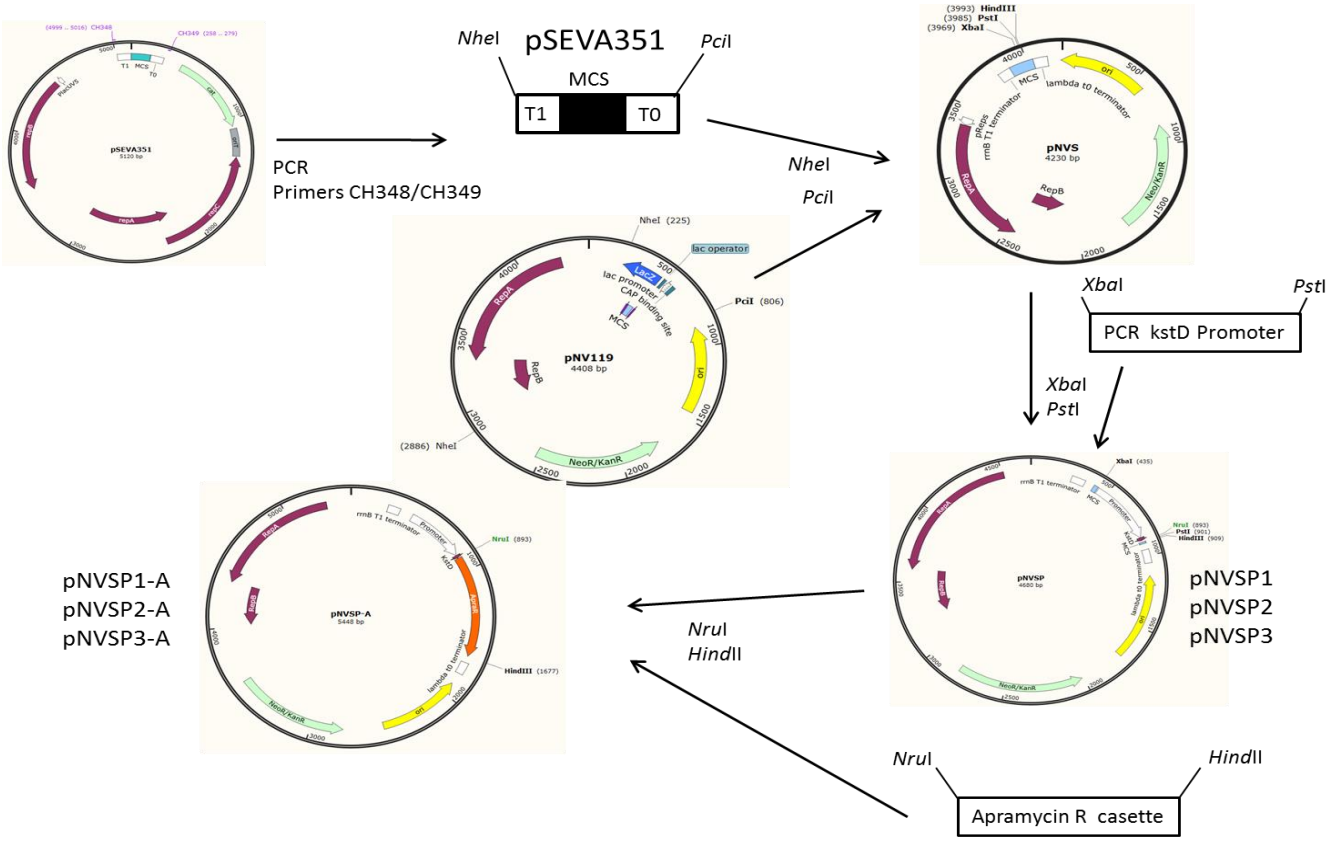
enzyme different to KstD1 and KstD2 and suggest that it may be involved in AD-independent steroid catabolism. The differences found in the orientation of catalytic residues of each KstD within the binding pocket site could explain the substrates preferences of each enzyme.

From the growth experiments, we can conclude that KstD2 and KstD3 are related to the long chain sterols metabolism as it was previously reported (Fernández de las Heras, van der Geize *et al.* 2012). KstD2 is also indispensable for AD growth but that seems not to be the case for testosterone and DHEA; when testosterone is used as carbon source, even the triple *kstD R. ruber* mutant is somehow able to grow. However, full-growth on this substrate is only reached when KstD2 is present in the cell. By the other hand, the growth in DHEA seems to be KstD1,2,3-independent. From these results we could suggest that there is other KstDs or similar enzymes involved in the degradation of these compounds.

Paradoxically, although KstD1 is probed to be *in vitro* a functional 3-ketosteroid-Δ^1-dehydrogenase able to act on different steroids, no specific role could be assigned to this enzyme *in vivo*. Potential functions of *R. ruber* strain Chol-4 of this isoenzyme and the other KstDs in other steroid pathways remain to be elucidated.

The promoter regions that support transcription of *kstD* genes have been cloned and functionally identified. The three promoter boxes contain different expression patterns, from the TCCG motif found in the *kshA-kstD2* ORF intergenic region to the KstR boxes in the *hsaE-kstD3* intergenic region. Moreover, *kstD3* ORF was transcribed as a polycistronic *kstD3-hsd4B-choG* mRNA and was induced in cholesterol growing media, reinforcing the role of KstD3 in the cholesterol metabolism.

Supplemental material 1. Scheme of the pNVSP vectors construction.



The shuttle vector pNV119 was modified to include a mcs from pSEVA351 (pNVS) and after that coupled with the apramycin resistance ORF. Intergenic regions containing the putative *kstD* promoters were cloned in the mcs to obtain pNVSP vectors.

Supplemental material 2. *R. ruber* strain Chol-4 ORF4 and ORF5 protein sequences.

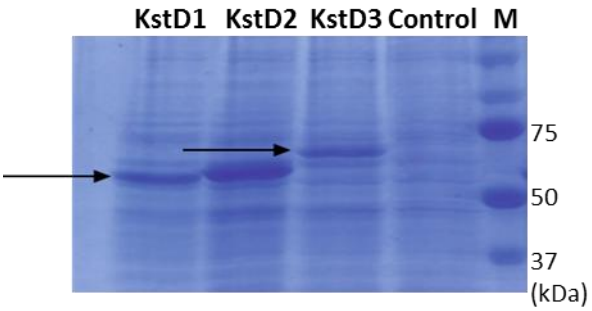
ORF4 (FAD-binding dehydrogenase)

MNPATVSESLTYDVVVVGSGAGGLSAAVAAAHGGASVLVVEKADVCGGATAWSGGWMWTPRNLFAHADGVHED
RAQPRRYLEHRLGDEFDAAKVDAFLDGAPEMVEFFERHTALQFVPGAAIADIHGDTPGAGTGHRSVAPKPVSLRRLGS
DVAALLRRQLYETSFLGMGIMAGPDLQAFLHSTRSGRAFVHCARRVSTHMLDLATERRGQQLVNGTALVGRLLRSALD
AGVDIRVATAATALVTDPSGRVTGVRIDGPGGARTVSARRGVVLATGGFTHDIDRRRELFPRTPSGREHWTLTPPTTTG
DGISLGESVGGRLDRSLASPVAYCPVSLVRYRNGRQGVFPHILDRGKPGVIGVLADGRRFVNEALGYHDYTLAMIEQVP
DGEEVCSWLIADQQYLRYFPLGMAKPFPIPTWPYLRSGYLSKGRTIRDLAEKIGVDPDGLEKTVTAFNESARVGEDPEFG
RGTTPFNVKSGDADNPWPNPSLAPLERGPFYAVKVVPGSFGTFAGLVTDSSSRVLNGDDRPIDGLFAVGVDQSSVMG
GHYPSGGINLGPAMTFGYLTGRRLASTTGATR

ORF5 (Fumarate reductase)

MQHNEEAAEYDVVVLGSGAAGLCAALSAARSGARVGVFEKGELLGGTTCLSSAVAWLPNNRYAREAGIADSREGALA
YLESLSHGMILPELAEAFVDTVPELLEWLDTTPLKMRLVAGYPDYHPERPGGMPHGGRSLEPELFSFIDLGRWEDKLVG
VPRRMTVTETPIGGGTGHLPADVQEQRERDHVEGLGRGMVAALLQGCLDEGVAVHTGERGVRLIQDESGRVTGVRF
EGRGGPHDVLAEHGVVLATGGFEWDERLRRDFLRGPLAHPATVPTSTGDGLRMAMRVGAQLGNMREAWWAPVA
VLPGQRANGAQAVQLVHRERTAPHSIMVNRHGRRFTNEATNYNALGGAFHHLDAHDFDYPNQPCWLIFDADHVEK
YGAFGAAPGTEAPEWVVRADTLADLARQIEVPAAALESTVAHWNEDVRRGHDSEYHRGESVYDGFVGDKNKYPGVE
STLGPVARAPFHAVQIHSSTLGTKGGPRTDADGAVLDVDDRVIPGLFAAGNVMAAPTGMVYGGAGGTLGPALVFGY
RAGRAAARAAGSPSEPGGVLREPTTV

Supplemental material 3. Expression of KstDs from ptip-QC1 vectors in induced *R. erythropolis* CECT3014 cells. SDS-PAGE analysis on a 12.5% gel was performed using 10 µg of the cell-free extracts. Control: *R. erythropolis* CECT3014 cells harbouring an empty ptip-QC1 vector. M: marker protein size.



CAPÍTULO 3

Characterization of 3-Ketosteroid 9α-Hydroxylases in *Rhodococcus ruber* strain Chol-4

1. Background

Steroid catabolism is widely spread in Actinobacteria (*Mycobacterium* and *Rhodococcus spp.*) and Proteobacteria (*Comamonas* and *Pseudomonas spp.*) (Bergstrand, Cardenas *et al.* 2016). Besides the general interest in knowing the details of the complex steroid degradation pathways, steroid intermediaries of high economic importance or strong biological activity are produced during the catabolic process (Yam and Okamoto 2011; Bhatti and Khera 2012; Donova and Egorova 2012; Yao, Xu *et al.* 2014), which is appealing enough to attract biotechnological interests and a great deal of research efforts.

Cholesterol is one of the steroids that have concentrated more studies (van der Geize, Yam *et al.* 2007; García, Uhía *et al.* 2012). The proposed catabolic oxidative pathway is formed by a complex set of enzymatic reactions. The first of them are included in the so-called upper pathway, and are: the oxidation of cholesterol to 4-cholesten-3-one, the carbon side chain cleavage at C17, in a similar way to the β-oxidation of fatty acids, and the polycyclic ring opening (Nesbitt, Yang *et al.* 2010; García, Uhía *et al.* 2012; García-Fernández, Frank *et al.* 2013).

The cholesterol catabolism combines several activities such as cytochromes P450 (Donova and Egorova 2012; McLean, Hans *et al.* 2012), cholesterol oxidases (Doukyu 2009; Vrielink and Ghisla 2009; Fernández de las Heras, Perera *et al.* 2014), dehydrogenases (e.g. KstDs) (van der Geize, Hessels *et al.* 2000; van der Geize, Hessels *et al.* 2002; Fernández de las Heras, van der Geize *et al.* 2012), or monooxygenases (e.g. the two component Rieske non-heme monooxygenase KshAB or the flavin-dependent monooxygenase HsaAB) (Dresen, Lin *et al.* 2010; Petrusma, Hessels *et al.* 2011; Petrusma, Dijkhuizen *et al.* 2012). The complexity of the cholesterol catabolic route lies on the facts that these steps are not always consecutive, that there is a multiplicity of enzymes with the same activity and that different aerobic pathways occur in a strain-dependent way. For instance, although there are a lot of oxygenases involved in this pathway, it has been recently described an aerobic cholesterol degradation pathway that it is oxygenase-non dependant (Wang, Lee *et al.* 2013). Therefore, this catabolism can be very versatile and a particular compound, ADD for instance, may be produced as a result of different activity sets (Yeh, Kuo *et al.* 2014).

KshAB complex is responsible for the C9α-hydroxylation; it consists of a terminal oxygenase (KshA subunit) which performs the substrate hydroxylation and a ferredoxin reductase (KshB subunit) which mediates the electron transfer. This enzyme initiates the steroid polycyclic ring opening in the steroid catabolism in combination with the KstD activity (Petrusma, van der Geize

et al. 2014). KshA contains two domains: a N- terminal Rieske iron-sulfur domain and a C-terminal catalytic domain that utilizes a non-heme Fe(II) to catalyse the addition of one hydroxyl group to the aromatic ring. KshB is similar to the ferredoxin reductase of multicomponent oxygenases; it is composed of a NAD(P) binding domain in the N-terminal region and a 2Fe-2S iron-sulfur binding domain in the C-terminal region. The reaction depends on the use of an electron transport chain that transfers electrons from the NADH to KshB and then to KshA, which has an O_2 bounded needed for the hydroxylation process of the substrate (Petrusma, van der Geize *et al.* 2014). The catalytic mechanism of KshAB is not yet clarified but it follows a pattern similar to other Rieske non-heme oxigenases: the substrate binds to the active site and the non-heme iron is reduced; when the oxygenase is in a reduced state, O_2 can bind to the metal centre and activate the hydroxylation process (Barry and Challis 2013; Petrusma, van der Geize *et al.* 2014).

Particularly, AD is substrate for both KstD and KshAB enzymes that yield ADD and 9OH-AD respectively; these products are further transformed with the same combination of enzymes to finally yield the unstable compound 9OH-ADD that spontaneously -by the break of the B ring- transforms into 3-HSA and enters into the so-called medium pathway that finally leads into the primary metabolism via the tricarboxylic acid cycle (Pandey and Sassetti 2008). The above scheme keeps ADD concentration at a low level avoiding cytotoxic effects (van der Geize, Hessels *et al.* 2008). The reactions starting from ADD, an intermediate from the degradation of several steroid compounds (Horinouchi, Kurita *et al.* 2010), are also known as 9,10-seco-pathway.

Among *Rhodococcus* strains, several KstD and KshAB isoenzymes have been reported. The wide enzymatic versatility and catabolic capacities of these activities make difficult to get a complete and detailed map of this catabolism. *Rhodococcus* strains are not yet biotechnologically fully exploited being the knowledge of all the steps of the steroid degradation pathways a serious bottleneck. Only recently, a *Rhodococcus erythropolis* strain got through a suitable combination of mutations in these enzymes has been used as whole-cell industrial steroid biotransformer (Venkataraman, Te Poele *et al.* 2015). In this work we identified and characterized three different KshA isoforms in *R. ruber* Chol-4 (Fernández de las Heras, García Fernández *et al.* 2009; Fernández de las Heras, van der Geize *et al.* 2012), a necessary step for broadening its biotechnological applications.

2. Results and Discussion

2.1. *In silico* analysis and identification of the putative *kshA* and *kshB* ORFs of *R. ruber* strain Chol-4.

3-ketosteroid 9α-hydroxylase (KshAB) and 3-ketosteroid-Δ^1-deydrogenase (KstD) are good examples of the high enzymatic redundancy that occur in Rhodococci. In particular, a deep search in the NCBI database led to find several strains that contain a multiplicity of KshA proteins: there are at least three putative isoforms of KshA in *R. pyridinivorans* KG-16 (AZXY01000000); four KshA subunits in *R. erythropolis* PR4*, R. qingshengii* BKS 20-40 and *R. aetherivorans* strain BCP1 (AP008957; NZ_AODN00000000; NC_008269; AVAE01000000); five KshAs in *R. rhodochrous* and *R. jostii* RHA1 (Petrusma, Hessels *et al.* 2011) and seven KshAs in *R. equi* NBRC 101255=C7 (BCRL00000000).

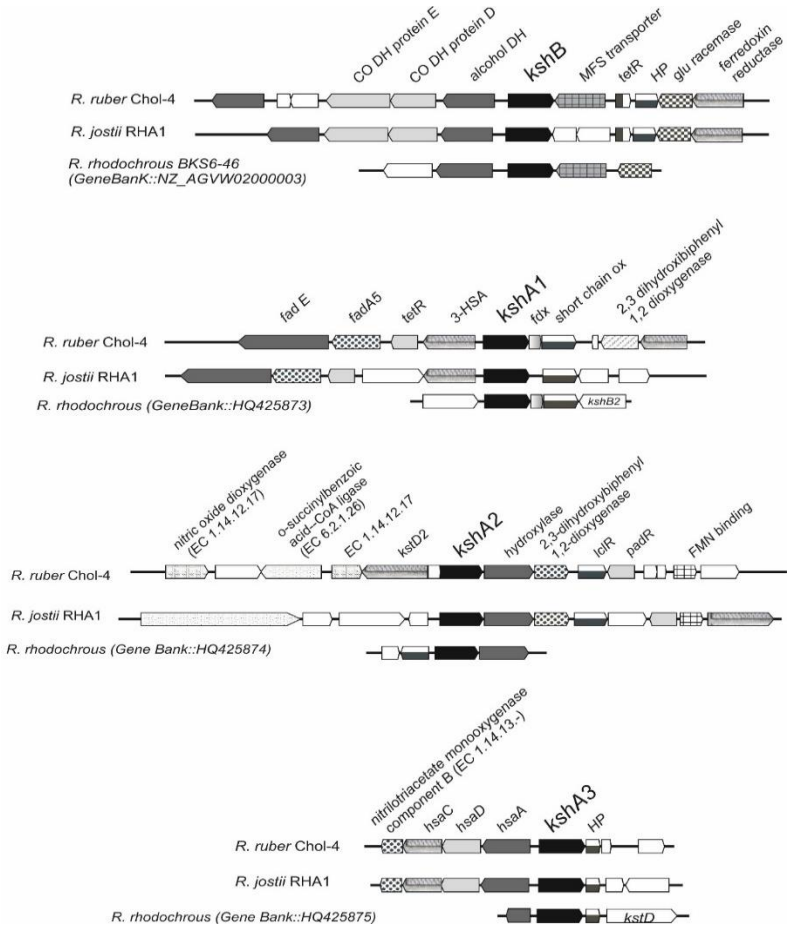


Figure 1. Scheme of *R. ruber kshAB* genomic contexts. A comparison with *R. jostii* Rha1 (GeneBank: NC_008268.1) and *R. rhodochrous* gene homologues is included. Abbreviations: *kstD*: 3-ketosteroid Δ^1-dehydrogenase; *hsaC*: 2,3-dehydroxyphenyl dioxygenase; *hsaD*: 4,5:9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-diene-4-oate hydrolase; *hsaA*: 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione monooxygenase, subunit A; *kshA*: 3-ketosteroid 9α-hydroxylase; *tetR*, *padR*, *lclR*: transcriptional regulators; HP: hypothetical protein; *fadA5*: acetyl-CoA acetyltransferase; *fadE*: acyl-CoA dehydrogenase; *ox*: oxygenase; *CO DH*: Carbon monoxide dehydrogenase.

The three KshAs (GeneBank::ACD11366, BAH31462, AAL96829) identified in *R. erythropolis* SQ1 (van der Geize, Hessels *et al.* 2008) and the five isoforms (GeneBank:: KshA1: ADY18310, KshA2: ADY181316, KshA3: ADY18318, KshA4: ADY18323, KshA5: ADY18328) of *R. rhodochrous* (Petrusma, Hessels *et al.* 2011) were blasted against the *R. ruber* strain Chol-4 genome (Fernández de las Heras, Alonso *et al.* 2013). Three different *kshA* ORFs were found in different genomic contexts (Figure 1). They were named as *kshA1* (GeneBank::KXF8587), *kshA2* (GeneBank::KXF84594) and *kshA3* (GeneBank::KXF84022) on the basis of their similarities to the *R. rhodochrous* isoforms.

Supplemental Table S1 collects the identity data of the protein coded by these ORFs and their genomic surroundings. The cladogram (Figure 2) shows that all KshAs could be grouped into three branches. KshA1 and KshA3 of *R. ruber* are in the first two branches and show a 97% and a 95% of amino acid (aa) identity when compared to the homologues of *R. aetherivorans* GenBank::AKE91471 and AKE88516, respectively. In contrast, *R. ruber* KshA2 has a relatively low aa identity with the rest of KshA proteins. The three *R. ruber* KshAs show a 63-65% aa identity among them.

The different rhodococci known genomes code for a less variable number of KshB copies: one isoform in *R. erythropolis* PR4 (AP008957) and *R. equi* NBRC 101255=C7 (BCRL00000000); two isoforms in *R. rhodochrous (Petrusma, Hessels et al. 2011)* and *R. opacus* (AP011115.1); and three isoforms in *R. jostii* RHA1, although one of them is associated to a plasmid and presents a higher divergence (Figure 2). The KshB proteins from *R. erythropolis*, *R. rhodochrous* and *R. jostti* RHA1 (van der Geize, Hessels *et al.* 2008; Mathieu, Mohn *et al.* 2010; Petrusma, Hessels *et al.* 2011; Mohn, Wilbrink *et al.* 2012) were used to blast against the *R. ruber* Chol-4 genome. Only one *kshB* ORF could be found (GeneBank::KXF85152) (Figure 1) that keeps a 99% of aa identity with one homologue of *R. rhodochrous* (see Supplemental Table S1). Its genomic location is not near any KshA, similarly to the cases of *Mycobacterium* and other *Rhodococcus* (van der Geize, Hessels *et al.* 2002; van der Geize, Yam *et al.* 2007). The lesser multiplicity of KshB respect to KshA subunits may be due to its role in the electron transfer to the KshA subunit: the same KshB molecule can work with several different KshA subunits whose catalytic domains will finally define the substrate specificity. In fact, the same KshB reductase was functional with KshA1 and KshA2 subunits from *R. erythropolis* SQ1 (van der Geize, Hessels *et al.* 2008); one KshB reductase was positively coupled to the five KshA homologues of *R. rhodochrous* DSM43269 (Petrusma, Hessels *et al.* 2011); and a quimeric *Mycobacterium* KshA – *Gordonia* KshB was proved to be

functional (Yuan, Chen *et al.* 2015). Furthermore, there are evidences that KshB can also act as a reductase for other oxygenases as well as for KshAs (Hu, van der Geize *et al.* 2010).

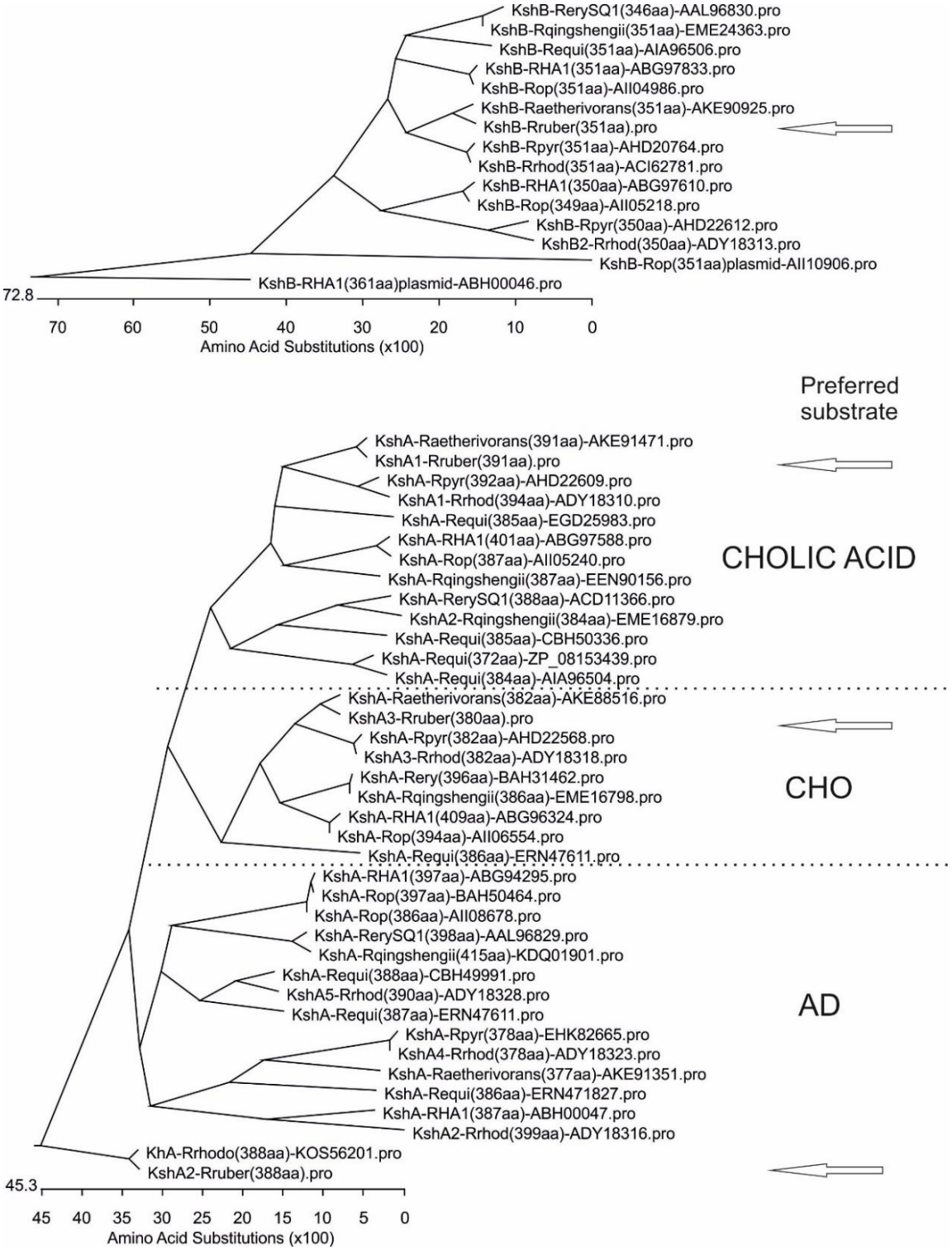


Figure 2. Cladogram of *Rhodococcus* KshBs and KshAs. Putative substrate preference of KshAs is indicated.

Multiplicity of these steroid catabolic enzymes seems to be a general characteristic of actinobacteria, as *Nocardioides simplex* VKM Ac-2033D contains in its genome two genes orthologous to *kshA* and two genes orthologous to *kshB* (Shtratnikova, Schelkunov *et al.* 2016) and *Mycobacterium abscessus* contains at least two *kshA* ORFs (NCIB::EIV82964 and EIV71174).

2.2. Molecular modelling of KshAs of *R. ruber* strain Chol-4

In an attempt to go deeper inside the characterization of these enzymes, we performed protein sequence analysis and modelling studies on the *R. ruber* KshAs using different approaches. The published three-dimensional structure and the catalytic mechanism proposed for KshA from *Mycobacterium tuberculosis* and KshA5 from *R. rhodochrous* have been used as suitable models (Capyk, D'Angelo *et al.* 2009; Penfield, Worrall *et al.* 2014).

The three KshAs found in *R. ruber* contain the Rieske Fe-S binding motif DAY**C**R**H**MGGdLsQGeiKGDSVA**C**PF**H**DWR and the non-heme iron binding motif EIi**D***NvV**D**MA**H**FfYv**H**Ysf reported for other KshA homologues: the consensus among the three *R. ruber* KshAs are written in capital letters (Petrusma, van der Geize *et al.* 2014); the asterisk marks the conserved aspartic acid Asp^{178} that is involved in catalysis and that also facilitates the electron transfer between adjacent subunits not only in KshAB but also in similar enzymes such as phthalate dioxygenase (Parales, Parales *et al.* 1999; Pinto, Tarasev *et al.* 2006; Tarasev, Pinto *et al.* 2006; Capyk, D'Angelo *et al.* 2009); in *Mycobacterium tuberculosis* there are two histidines (His^{181}, His ^{186}) and a Asp^{304} that are involved in the iron coordination at the core of the catalytic domain (Capyk, D'Angelo *et al.* 2009).

Similarly, KshB from *R. ruber* contains the flavin binding motif SVA**R**C**YSL**ASS, the NADH binding motif LFGA**G**S**GITP**VVS and the plant-type Fe_2-S_2 binding motif PYS**C**QEGECGS**C**A**C**TVLDGKVEME HCEILDPEDIEAGYVLG**C**QAR. In both KshA and KshB regions, the conserved residues in actinobacteria described by Petrusma *et al.* (Petrusma, van der Geize *et al.* 2014) appear in black and underlined.

I-TASSER and PredictProtein programmes predicted that none of the three *R. ruber* KshAs contain sulfur bridges or transmembrane segments. The prediction of the secondary structure and solvent accessibility of the three KshAs from *R. ruber* are shown in Supplemental Table S2. Table 1 collects the described residues involved in the active site of *R. rhodochrous* KshA5 (Penfield, Worrall *et al.* 2014) and the homologous residues of the three KshAs of *R. ruber* found by COBALT programme.

The eight key residues (Val^{182}, Gln^{210}, Tyr^{239}, Met^{245}, Asp^{247}, Asn^{264}, Phe^{308} and Trp^{315}) of the *R. rhodochrous* KshA5 active site have counterpart residues in the *R. ruber* KshAs. However, the I-TASSER model prediction of these *R. ruber* enzymes shows that the orientation of the key residues within the catalytic pocket differs, being specific of each enzyme (Figure 3). Particularly, the orientation of KshA1 Tyr^{233} benzene ring differs in about 45 degrees from that of its KshA2 and KshA3 homologues, a difference that may cause an impact in the anisotropic properties of the catalytic pocket. The comparison of the relative position of several homologous residues of the three KshAs *R. ruber* models has showed differences that may be meaningful. This is the case of F^{301}, M^{239}, Y^{233}, W^{308} Ksha1 residues and their respective KshA2 and KshA3 homologous residues (see Table 1).

Table 1: Homologous amino acid residues implicated in the catalytic pocket of KshAs. Grey lines contain the key residues of the KshA5 active site of *R. rhodochrous* (Rrho) and its homologues in *R. ruber* given by COBALT.

Rrho	KshA1	KshA2	KshA3
I^{178}	I^{172}	I^{176}	I^{166}
N^{181}	N^{175}	N^{179}	N^{169}
V^{182}	I^{176}	V^{180}	V^{170}
H^{187}	H^{181}	H^{185}	H^{175}
F^{188}	F^{182}	F^{186}	F^{176}
H^{192}	H^{186}	H^{190}	H^{180}
S^{194}	S^{188}	S^{192}	A^{182}
Q^{210}	Q^{204}	Q^{208}	Q^{198}
M^{212}	M^{206}	M^{210}	L^{200}
S^{214}	G^{208}	F^{212}	T^{202}
L^{233}	S^{227}	-	L^{220}
S^{235}	S^{229}	S^{232}	S^{222}
A^{237}	A^{231}	A^{234}	A^{224}
Y^{239}	Y^{233}	Y^{236}	Y^{226}
M^{245}	M^{239}	M^{242}	M^{232}
D^{247}	D^{241}	D^{244}	N^{234}
L^{249}	L^{243}	L^{246}	L^{236}
L^{262}	L^{256}	L^{259}	L^{249}
N^{264}	N^{258}	N^{261}	N^{251}
V^{304}	I^{297}	T^{301}	I^{291}
F^{308}	F^{301}	F^{305}	F^{295}
D^{311}	D^{304}	D^{308}	D^{298}
W^{315}	W^{308}	W^{312}	W^{302}

Moreover, KshA1 has an Ile$_{176}$ where its KshA2 and KshA3 homologues have a Val. Possibly, some or most of these differences will finally be related to the different substrate preference among these KshAs. At the moment, the current shortage of crystalline structures of these proteins greatly limits more detailed conclusions.

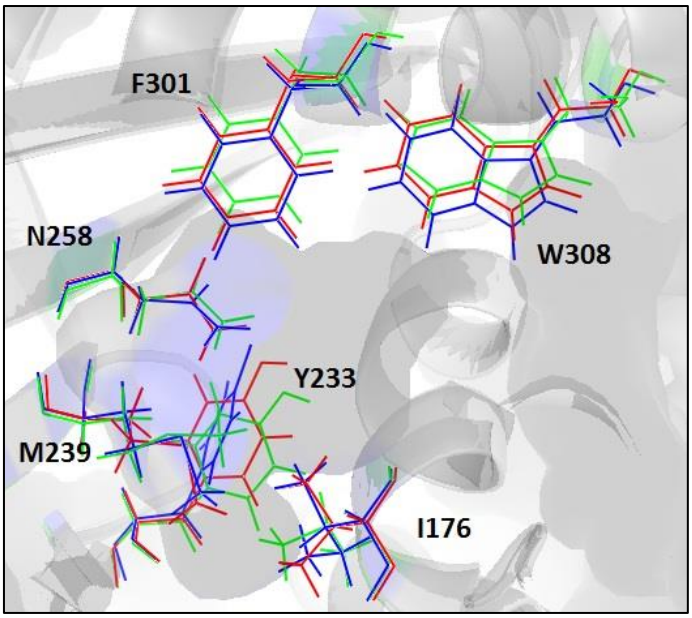


Figure 3. Modelling of the active site of KshAs. The orientation of some of the catalytic residues in the KshA binding pocket is shown. They are superposed and depicted in different colors: blue for KshA1, red for KshA2 and green for KshA3. Only KshA1 residues are named, their counterpart homologues residues of KshA2 and KshA3 being listed in Table 1.

2.3. *KshAB* ORFs co-transcription studies

The *R. ruber kshAB* genes surroundings contain ORFs that could be also involved in the steroid metabolism (Figure 1). Some of these ORFs may be expressed in a coordinated way or even co-transcribed with *kshA* genes. RT-PCR confirm that all *R. ruber kshA* genes are transcribed together with adjacent ORFs giving rise to long transcripts (*kshB* gene is flanked by opposite ORFs preventing co-transcription).
kshA1 ORF is co-expressed with an hypothetical protein (Fdx) that keeps a 58% of amino acid identity with a ferredoxin (ADY18311.1) of *R. rhodochrous* and with an ORF that has 98% aa identity with a short chain dehydrogenase of *R. rhodochrous* (Table S1). The cluster *kshA1-fdx*-short chain dehydrogenase might be conserved within the *Rhodococcus* genus (Figure 1).

kshA2 ORF is co-expressed with an ORF that shows a 98% of aa identity with a hydroxylase of *R. rhodochrous* (WP_054372665), another ORF 94% similar to an 2,3-dihydroxybiphenyl 1,2-dioxygenase of *R. rhodochrous* (WP_054372664) and a third ORF 95% similar to an IclR family transcriptional regulator (*R. rhodochrous*, WP_054372663) (Table S1). This gene cluster organization is identical to that of *R. jostii* Rha1 but something different to that of *R. rhodochrous* BKS6-46 (GeneBank::NZ_AGVW02000046) (Figure 1).

Lastly, *kshA3* ORF is co-expressed with two ORFs that may code for two hypothetical proteins that do not contain conserved domains. The first ORF keeps a 63% of aa identity with a putative

ferredoxin (ADY18319) of *R. rhodochrous.* The second ORF has a 99% aa identity with a putative protein (WP_054371650) of *R. rhodochrous*. The putative ferredoxin is conserved in both *R. jostii* Rha1 and *R. rhodochrous* but the second ORF is different in each strain. The presence of a putative ferredoxin near a *kshA* ORF could be related to the proposed role of ferredoxin as an enhancer of the electron transfer between KshA and KshB (Petrusma, Hessels *et al.* 2011).

2.4. Transcriptional analysis of *kshA* genes in *R. ruber* strain Chol-4

The transcription of the three *kshA* genes found in *R. ruber* strain Chol-4 was analysed by RT-qPCR. RNA samples were prepared from cultures grown in M457 mineral medium supplemented with AD and cholesterol as possible inductors. Control cultures were grown in M457 supplemented with sodium acetate (2 g/L NaAc) in the absence of any steroid. Using the specific primer pairs designed to search for the transcript of each different ORF (see Materials and Methods, Table 6), we could see that all the *kshA* genes were transcribed in all the conditions used in our assays. The expression levels on sodium acetate (unexposed steroid culture) for the reference gene (*D092_14375)* was taken as 1. The preliminary values obtained for the expression of the three *ksh* genes were: 8, 248 and 198-fold change for *kshA1*, *kshA2* and *kshA3*, respectively, in cultures grown in cholesterol; only *kshA2* was overexpressed (5702-fold change) in cultures grown in AD.

So we can conclude that these three genes, although basically transcribed, can be induced by the presence of cholesterol (*kshA1*, *kshA2* and *kshA3*) or AD (*kshA2*). The induction of some of the isoforms of *kshA* by cholesterol or AD has also been reported by conventional RT-PCR in other *Rhodococcus* microorganisms such as in *R. erythropolis* or *R. rhodochrous* (van der Geize, Hessels *et al.* 2008, Petrusma, Hessels *et al.* 2011).

Up to five *kshA* homologues (*kshA1* to *kshA5*) have been previously identified in *R. rhodochrous* DSM43269 showing all of them a different expression pattern: *KshA1* is induced in cholic cultures and faintly in cholesterol; *kshA2* is induced in AD and cholic cultures; *kshA3* is hardly expressed in any condition and faintly in AD; *kshA4* is induced in cholesterol and *kshA5* is induced in progesterone (Petrusma, Hessels *et al.* 2011). The *kshA1* gen of *R. rhodochrous* had the same expression patter than its homologue in *R. ruber kshA1*, in the presence of cholesterol it is lightly induced and it is not induced in AD. On the other hand, although *kshA3* genes of *R. ruber* and *R. rhodochrous* share the highest identity, their expression patterns are not the same; while *kshA3* of *R. rhodochrous* seems to be induced by AD in *R. ruber* there is not sign of induction of the *kshA3* isoform (Petrusma, Hessels *et al.* 2011). Respect to *kshA2* of *R. ruber*, it seems that this

isoenzyme does not share a defined identity with any of the five *kshAs* of *R. rhodochrous*, although it could share some common characteristics with *kshA2, ksh4* or *ksh5.* Therefore, due to the fact that the number of *kshA* isoenzymes and that the regulation pattern differ within *Rhodococcus* strains, the catabolism of steroids could have developed from a common core to a specific scenario strain-dependent.

2.5. Construction of *R. ruber* strain Chol-4 mutants

KshAB plays a role at different steps of the steroids degradation pathway, as they can act on a variety of substrates with several A-steroid ring configurations and also with different C17 side chains (Capyk, Casabon *et al.* 2011; Petrusma, Hessels *et al.* 2011; Petrusma, van der Geize *et al.* 2014). Only few studies on the physiological roles of the *kshA* isoforms have been reported to date. In this piece of work we have made *ΔkshA* and Δ*kshB R. ruber* strains and studied their growth on different substrates in order to clarify the role of every isoform in this complex metabolic pathway.

Once identified the putative *kshA* and *kshB* genes in the genome of *R. ruber*, we have proceeded to build single, double and triple *kshA* mutants and the single *kshB* mutant of this strain following the unmarked gene deletion procedure described before. Care has been taken to keep the reading frame after truncation of the *kshA* and *kshB* genes, in order to avoid possible polar effects in the expression of the *kshA* and *kshB* operons. Single mutants (*ΔkshA1, ΔkshA2, ΔkshA3, ΔkshB)*, double mutants *(ΔkshA1,2, ΔkshA2,3, ΔkshA1,3)* and the triple mutant *ΔkshA1,2,3* were obtained for this work. None of these mutants were impaired to growth in rich medium or showed growth differences in minimal medium supplemented with sodium acetate.

2.6. Growth of *R. ruber* strain Chol-4 *ΔkshA* and *ΔkshB* strains in different media

The capability of the *R. ruber* mutants to grow on different steroid media was checked by cultivating them in 457 DSMZ minimal medium supplemented with the desired carbon source (Supplemental Figure S3). Steroids were prepared in 16 mM cyclodextrins as indicated in methods (*R. ruber* was not able to grow in cyclodextrins as single carbon source). Table 2 shows the data obtained after a 48 h growth of the wild type and the mutant strain cultures.

These results show that KshA1 is the only KshAB enzyme involved in the cholic acid catabolism while it also participates in the degradation of other substrates used in this study. On the other hand, KshA2 is mainly involved in the catabolism of AD, ADD, testosterone and DHEA, while KshA3 is exclusively related to the cholesterol and phytosterol degradation.

Table 2: Growth of *R. ruber* wild type (WT) and *kshAB* mutants on different substrates after 48 hours of culture.

Carbon source/ Strain	CHO 1.6 mM	AD 2.2 mM	ADD 2.2 mM	Cholic 2.2 mM	Tes 2.2 mM	DHEA 2.2 mM	Phy 0.6 mg/ml	Acet 2 mM
ΔkshA1,2	++	-	-	-	-	-	++	++
ΔkshA1,3	+	++	++	-	++	++	-	++
ΔkshA2,3	++	+	+	++	-	++	++	++
ΔkshA1,2,3	+	-	-	-	-	-	-	++
ΔkshB	+	-	-	-	-	-	-	++
WT	++	++	++	++	++	++	++	++

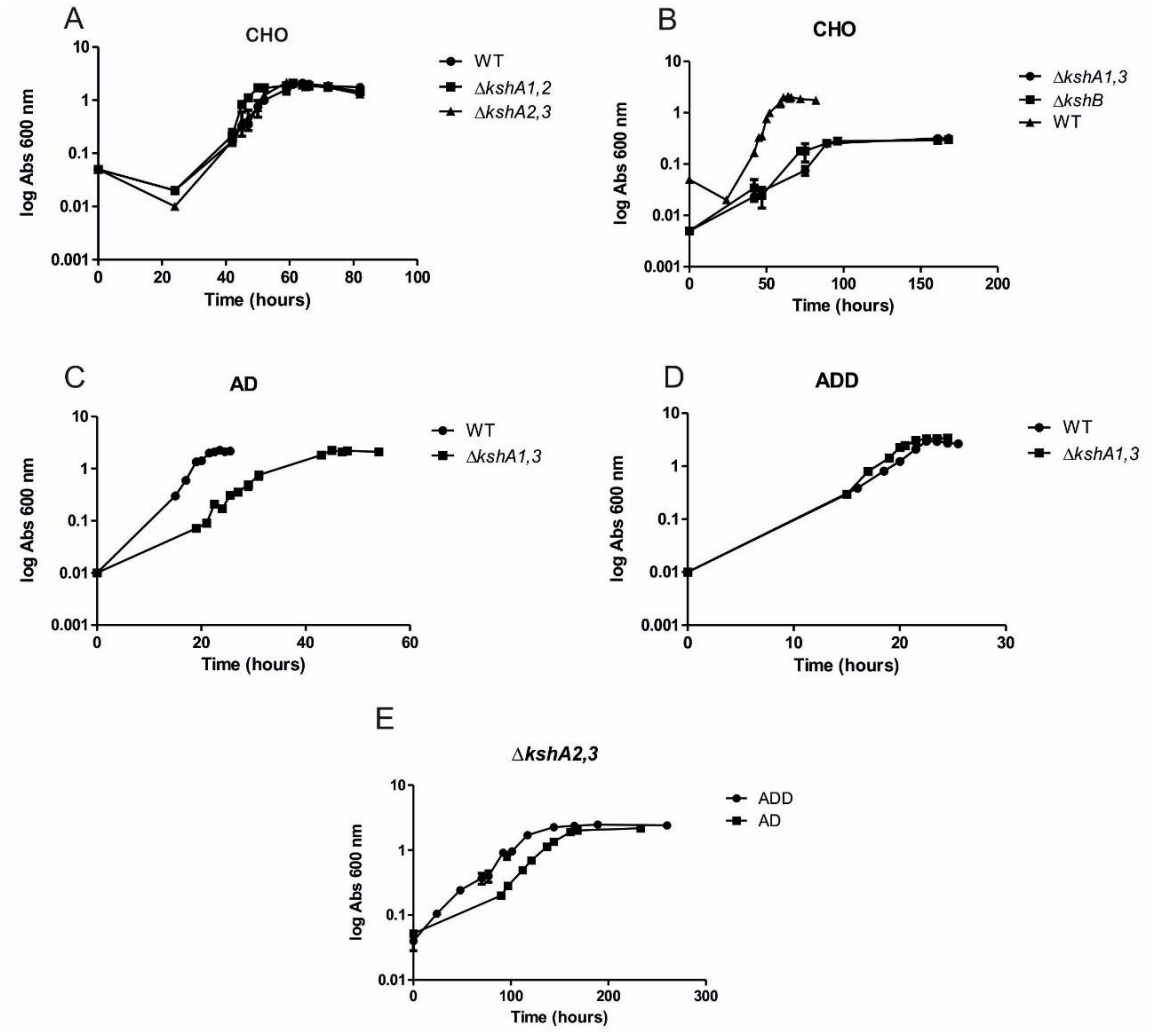
Cholic: cholic acid; **Tes**: testosterone; **DHEA**: dehydroepiandrosterone; **Phy**: phytosterols; **Acet**: sodium acetate.
+: A_{600nm} < 0.3; ++: A_{600nm} >0.3.

Growing behaviour of the *R. ruber* wild type and *kshAB* mutant strains on molecules most related to the cholesterol catabolic pathway was studied deeper. Growth curves of these strains are shown in Figure 4.

The growth parameters of the Δ*kshA2,3* strain growing in cholesterol are similar to those of the wild type (Figure 4), but its doubling time in AD or in ADD is 10 times that of the wild type (22 hours versus 2.2 hours) and the maximum absorbance at 600nm reached when growing in ADD is a 33% lower than the maximum got by the wild type (2.0 versus 3.0, Figure 4). It seems that KshA1 works very well in the cholesterol degradation pathway and rather worse in the AD or ADD catabolism. KshA3 shows a more strict behaviour, as the Δ*kshA1,2* strain can only grow on cholesterol and phytosterol substrates, but not in AD or ADD; its growing parameters on cholesterol are the same as those of the wild type strain, suggesting that KshA3 is mainly devoted to long chain steroid degradation. When KshA3 is lacking, KshA1 helps to do the work.

The presence of just the KshA2 activity in the Δ*kshA1,3 R. ruber* strain gave growth parameters in cholesterol similar to those obtained from the *R. ruber* Δ*kshA1,2,3* or from the *ΔkshB* strain (a maximum absorbance at 600 nm of 0.3 unit and more than 15 hours of doubling time), suggesting that cholesterol is not a direct substrate of the KshA2 activity. The small growth detected when some mutants grow in cholesterol can be attributed to the degradation of the side chain of the substrate molecule: growth can occur until the side chain is completely gone (see later).

On the other hand, the only presence in the cell of the KshA2 activity in the Δ*kshA1,3* strain is enough to obtain growth parameters on AD or ADD similar to those of the wild type. It can be concluded that the growth in AD and ADD is mainly mediated by KshA2 activity in *R. ruber* although KshA1 activity can also contribute as we have mentioned earlier.



Strain	KshA remaining	CHO 1.6 mM		ADD 2.2 mM		AD 2.2 mM	
		Amax 600 nm	DT (hours)	Amax 600 nm	DT (hours)	Amax 600 nm	DT (hours)
WT	KshA1,2,3	2.0 ± 0.1	4.7 ± 0.4	3.00 ± 0.10	2.2 ± 0.0	2.2 ± 0.0	2.1 ± 0.1
ΔkshA1,3	KshA2	0.3 ± 0.0	15.1 ± 0.7	3.3 ± 0.1	1.8 ± 0.1	2.1 ± 0.1	3.5 ± 0.4
ΔkshA2,3	KshA1	1.9 ± 0.2	3.9 ± 0.6	2.0 ± 0.5	21.9 ± 3.8	2.2 ± 0.0	19.8 ± 0.6
ΔkshA1,2	KshA3	2.1 ± 0.1	4.1 ± 0.6	ng	ng	ng	ng
ΔkshA1,2,3	none	0.3 ± 0.0	16.1 ± 1.2	ng	ng	ng	ng
ΔkshB	KshA1,2,3	0.3 ± 0.0	18.9 ± 1.3	ng	ng	ng	ng

Figure 4. Growth curves of *R. ruber* wild type (WT) and mutant strains in minimal medium supplemented with different steroids as only carbon source. A representative experiment of 3-4 replicas is shown. The standard error of the mean was calculated by GraphPad Prism 5.0. The *ΔkshA1,2,3* strain growth curve on cholesterol of follows the same pattern as those of *ΔkshA1,3 or ΔkshB* strains in panel B. *ΔkshA1,2* strain did not grow in AD or ADD. Maximum absorbance at 600 nm and doubling time of cultures of *R. ruber* wild type (WT) and *kshAB* mutants on different substrates are shown in the table. ng: no growth.

R. ruber strains devoid of any KshAB activity (Δ*kshB* or Δ*kshA1,2,3* strains) were able to grow in cholesterol although 10 times slower than the wild type strain (Figure 4). An impaired growth on cholesterol has been also described for *R. rhodochrous* RG32 mutant (devoid of their five KshAs) (Petrusma, Hessels *et al.* 2011; Wilbrink, Petrusma *et al.* 2011).This growth is due to the degradation of the side chain, a process that has been described to occur at different steps of the pathway being independent of the polycyclic ring opening (Supplemental Figure S4) (Capyk, Casabon *et al.* 2011; Donova and Egorova 2012). Therefore, in these strains the side chain degradation process does not depend on any KshAB activity.

But there are some opposite evidences in other actinobacteria: a relationship between KshAB activity and the degradation of the side chain of cholesterol has been proposed in *R. equi* and *R. erythropolis* (van der Geize, Hessels *et al.* 2002; Yeh, Kuo *et al.* 2014). Even more, the physiological substrates of KshAB enzymes are described to be CoA thioester derivatives of cholesterol degradation intermediates with a long side chain in *Mycobacterium* (Capyk, Casabon *et al.* 2011) although this does not apply to all the *R. rhodochrous* KshA homologues (Penfield, Worrall *et al.* 2014). The dependence or not of the side chain degradation on the KshAB activities could be strain specific.

2.7. Genetic complementation of *R. ruber* strain Chol-4 mutants

R. ruber Δ*kshA1,2,3* , devoid of the three KshA activities, is totally unable to use many compounds as the only carbon and energy source (as AD and cholic acid) and is severely impaired to grow using CHO. This strain was successfully complemented with pNV119-derived pNVS recombinant plasmids harbouring a functional copy of *kshA1*, *kshA2* or *kshA3 R. ruber* genes. Figure 5 depicts how mutant cells transformed with the three pNVS-derived recombinant plasmids recovered or not the ability to grow in some particular steroid carbon sources, confirming the data of mutant growing experiments.

Complementation data of the *ΔkshA1,2,3* strain suggested that KshA2 is only able to act on substrates with a short side chain (AD, ADD, testosterone, DHEA) in *R. ruber*. On the other hand, KshA3 can only grow using cholesterol as carbon source.

KshA1 is essential for *R. ruber* to grow in cholic acid but it can also substitute KshA3 in the cholesterol degradation. KshA1 isoform may have a broader range of substrates as *R. ruber* Δ*kshA2,3* strain can also grow in ADD and AD although with a doubling time 10 times higher than when KshA2 is the only KshAB active in the cell. Surprisingly, complementation of the *ΔkshA1,2,3*

strain with KshA1 perfectly restores growth in CHO or cholic acid but not in AD, ADD or DHEA, whereas the *ΔkshA2,3* could do it to certain point (Figure 4 and 5). It could be possible that the growth on these substrates has some regulation requirements that the above complementation does not fulfil.

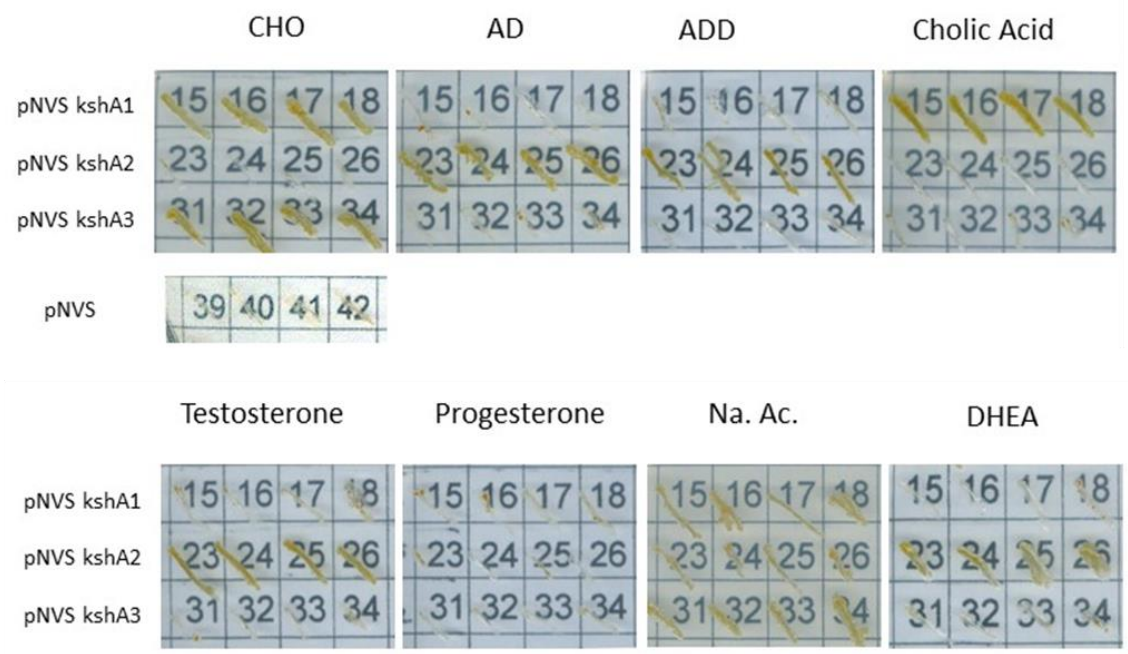


Figure 5. Complementation of the *ΔkshA1,2,3* strain with the pNVS vectors (pNV119-derived recombinant plasmids harbouring a functional copy of each *R. ruber kshA* ORFs).

Petrusma *et al.* proposed that each KshAB phylogenetic branch could contain enzymes functionally related with a similar spectrum of substrates (Petrusma, Hessels *et al.* 2011; Petrusma, van der Geize *et al.* 2014). This is confirmed by the involvement in the cholate metabolism of *R. ruber* KshA1, *R. rhodochrous* KshA1 (ADY18310) and *R. jostii* RHA1KshA (ABG97588), all of them belonging to the same branch (see Figure 2) (Mathieu, Mohn *et al.* 2010; Petrusma, Hessels *et al.* 2011; Wilbrink, Petrusma *et al.* 2011; Mohn, Wilbrink *et al.* 2012).

R. rhodochrous and *R. jostii* RHA1 contain up to five different KshAs and, consistently, the spectrum of their actions differs from that of *R. ruber* that only contains three KshAs. It has been reported in *R. rhod*ochrous that KshA1 is involved in the cholic acid metabolism; KshA2 is involved in the degradation of steroids with none or a short side chain; KshA3 is only involved in the CHO metabolism; KshA4 is involved in the degradation of steroids without any side chain; and lastly, KshA5 is a broad specificity isoform (Petrusma, Hessels *et al.* 2011).

Surprisingly, while the branches dedicated to cholic acid and CHO contain just one KshA isoform of *R. rhodochrous*, the branch related to AD metabolism (Figure 2) contains the rest of the isoforms of this bacterium. In general, the AD and cholate branches contain more KshA isoforms that the CHO branch (Figure 2).

R. ruber KshA2 is an isoform related to the AD, testosterone and DHEA metabolism. This isoform keeps only a 67% aa identity with a KshA from RHA1 (67% aa identity: ABG94295) but the genomic surroundings of both strains are similar (Figure 1).

Any of *R. ruber* KshA1 or KshA3 was able to restore the growth on cholesterol when expressed in the *R. ruber ΔkshA1,2,3* strain, while KshA2 restores the growth on AD. *R. ruber* KshA2 shows a substrate preference for steroids with a short side chain (AD, ADD). In *R. erythropolis* SQ1, KshA2 has been suggested to prevent the accumulation of ADD to reach toxic levels during sterol catabolism (van der Geize, Hessels *et al.* 2008). On the other side, KshA2 from *R. rhodochrous* was able to restore the growth on cholesterol in *R. rhodochrous* RG32 mutant (devoid of their five KshAs). The physiological roles of the KshA homologues are clearly strain-dependent.

2.8. Monitoring of degradation intermediaries by TLC

KstD and KshAB are two key enzymes involved in the steroid ring opening and therefore their gene deletions should block this opening to yield 9OH-AD or ADD, respectively (Supplemental Figure S4), two C19-steroid intermediaries that have drawn great attention from both the basic and applied (industrial) research (Fernandes, Cruz *et al.* 2003; Wang, Yao *et al.* 2011; Petrusma, van der Geize *et al.* 2014).

R. ruber wild type strain and *kshAB* knockout mutants grown on several steroid compounds were used to analyse possible catabolism intermediaries. The harvested resting cells were chloroform extracted and analysed by TLC as described in Materials and Methods.

Data from cultures in AD are shown in Figure 6 a, b. A similar accumulation of ADD was observed in cultures of all *kshAB* mutants, but not in the wild type cells. This result was the same in samples taken from both, cell cultures and resting cell assays. In cells with faint or null KshAB activity, KstD transformed AD into ADD which accumulated and was clearly detected (Figure 6 a, b).

This result is consistent with the data of culture growth of those *kshA* and *kshB* mutants that cannot grow on AD (*ΔkshB*, *ΔkshA1,2* and *ΔkshA1,2,3*) and even in the case of the *ΔkshA2,3* strain, able to grow in AD although with a very high doubling time (Figure 4). However, the ADD accumulation detected in cultures of the *ΔkshA1,3* strain, that grows in AD almost as well as the wild type does (Figure 4) was unexpected. KshA2 is the major activity in the AD/ADD hydroxylation process and seems to be somehow helped in this task by KshA1. The absence of KshA1 may reduce the production of hydroxylated molecules (what would lead to the accumulation of ADD) with no meaningful effect on the growing rate.

Two more conclusions can be drawn from these results: i) the ADD accumulation, at least to the level reached in these assays, is not a serious problem for the cell growth; and ii) the KstD activities work more efficiently than the KshA2B activity, as in samples of *ΔkshA1,3* cells growing in AD no trace of 9OH-AD is detected, while the accumulated ADD remained in the cell even after 48h (Figure 6 a). It should be noted that 9OH-AD was however detected in a *R. ruber kstD* mutant (Fernández de las Heras, van der Geize *et al.* 2012) and also when the chimeric KshAB (KshA from *Mycobacterium smegmatis* mc2 and KshB from *Gordonia neofelifaecis* NRRL B-59395) was heterologously expressed in the AD producer strain *Mycobacterium neoaurum* NRRL B-3805 (Yuan, Chen *et al.* 2015; Rodríguez-García, Fernández-Alegre *et al.* 2016).

TLC intermediary analysis in cells grown in cholesterol provide with other suggestions. In cultures of *R. rhodochrous* RG32 mutant (devoid of its five *kshA* genes) ADD and $\Delta^{1,4}$-BNC have been identified as intermediaries of the process (Petrusma, Hessels *et al.* 2011; Wilbrink, Petrusma *et al.* 2011). A similar result has been reported form *R. equi kshB* mutant (Yeh, Kuo *et al.* 2014). Our results with *R. ruber* Chol-4 Δ*kshA1,2,3* and Δ*kshA1,3* strains show that they can hardly grow on cholesterol (Figure 4), what we firstly explained by the low efficient degradation of the carbon side chain of the molecule, a process that therefore would be independent of the KshAB action. However, biotransformation assays with both the wild type and these *R. ruber* mutant cells growing in cholesterol yielded no significant traces of ADD or 9OH-AD spots in TLC (Figure 6 c). The absence of intermediaries of the cholate catabolism was reported after the GC-MS analysis of cultures of a *R. jostii* Rha1 *kshA* mutant (Mohn, Wilbrink *et al.* 2012).

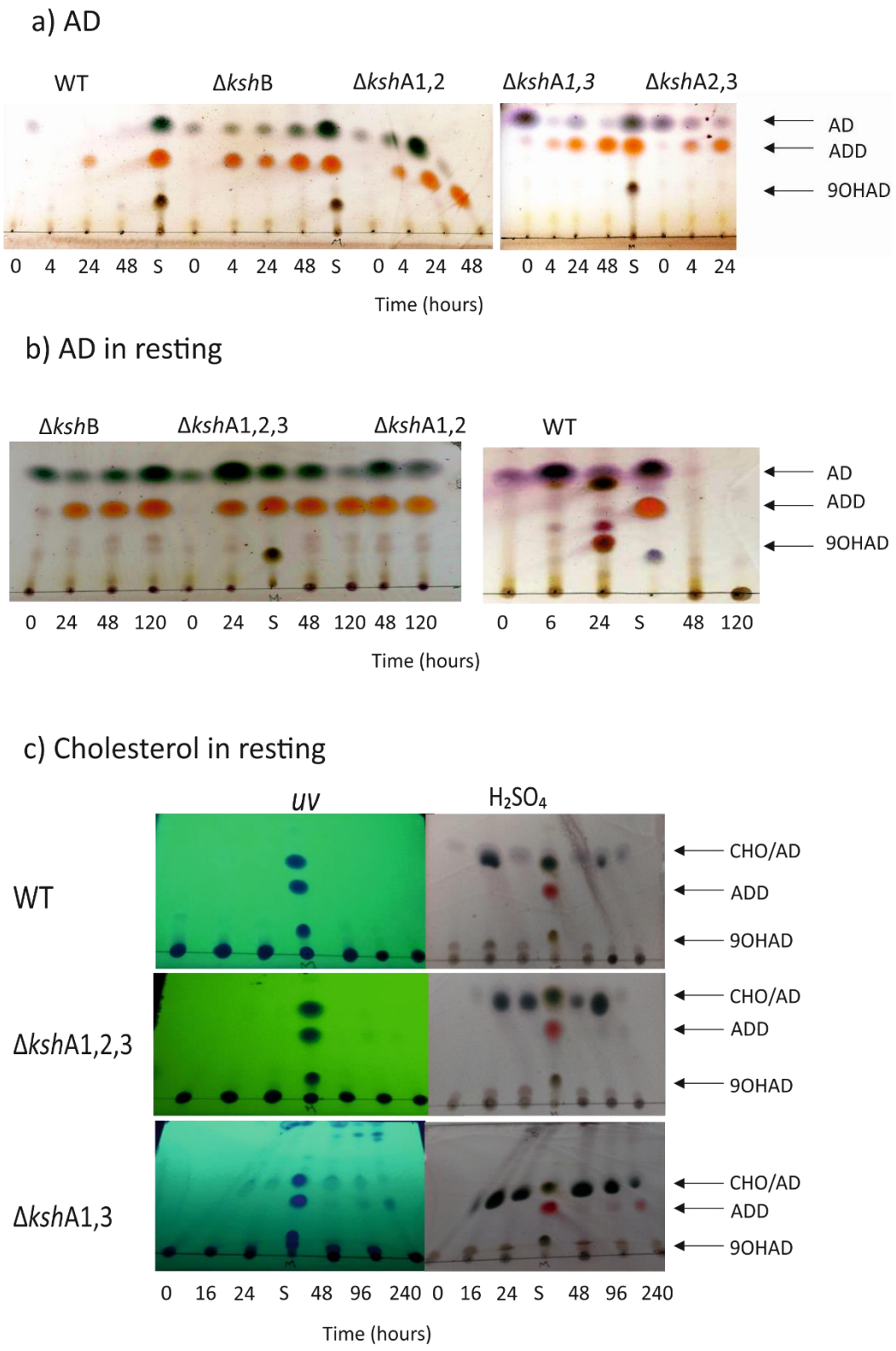


Figure 6. Thin-layer chromatography-based monitoring of steroid degradation intermediaries. A) TLC of samples from Chol-4 cultures growing in AD; b) and c). Biotransformation studies done on *R. ruber* strain Chol-4 WT and mutants on AD (panel b) and cholesterol (panel c).

3. Conclusions

These data lead to think that, owing to the different alternative combinations of isoforms, particular diversifications of the catabolic pathway may be possible among different strains. It has been recently proposed that sterols can be catabolised in *R. equi* USA-18 through two different routes (via AD or via $\Delta^{1,4}$-BNC) that converge in 9OH-ADD (Yeh, Kuo *et al.* 2014). Alternative pathways that do not pass through AD have been also suggested in the testosterone catabolism of *Steroidobacter denitrificans* (Chiang, Fang *et al.* 2010). In all those cases the degradation pathways would converge into ADD or 9OH-AD intermediaries, although a preference for steroid catabolism via 9OH-AD has been suggested in *Rhodococcus* (van der Geize, Hessels *et al.* 2008) and *Mycobacterium* (Yao, Xu *et al.* 2014).

In the light of the above opinions, the AD/ADD branch mediated by KshA2 in *R. ruber* Chol-4 may not be the main pathway of the cholesterol degradation in this strain. In this point attention must be focused on the KshA3 activity, whose only presence allows a good growing of *R. ruber* in cholesterol (Figure 6). KshA3-based pathway could be an alternative to the AD/ADD branch in the cholesterol catabolism of *R. ruber*, that would by-pass AD and ADD intermediaries. TLC spots are hardly visualised in cholesterol-grown resting cells only in the case of the *Δksh1,3* strain, where KshA3 is lacking and KshA2 is present (Figure 6 c). The absence of TLC spots when analysing cultures of wild type *R. ruber* grown in cholesterol suggests that the degradation of this molecule occurs mainly through the alternative KshA3 pathway. By contrast, AD/9OH-AD (both easily detectable by TLC) have been proved to be intermediaries from cholesterol degradation in a *kstD* mutant of *M. smegmatis* (Brzostek, Sliwinski *et al.* 2005), supporting the previous suggestion about different and particular alternative ways to catabolize cholesterol in bacteria.

In conclusion, in this study a collection of *kshA* and *kshB* mutants have been constructed in *R. ruber* and the role of these activities in the growth behaviour in different substrates have been discussed. KshA2 subunit is the main responsible of steroid substrates with short side chain while KshA3 subunit is specific for steroids with longer side chains. KshA1 activity, being necessary for cholic acid degradation, also helps to the other KshA3 and KshA2 isoforms in their action. These studies have also suggested that the side chain degradation during the steroid catabolism is independent of the KshAB activities in *R. ruber*. A cholesterol degradation pathway alternative to the AD/ADD branch in *R. ruber* and mainly based on the KshA3 and KshA1 activities is being subject of further studies.

Table S1: a) ORFs near to KshA1

ORF	aa/ pI/ kDa	Protein domains/Evalue	Max amino acid ID	*R. jostii*
3-HSA	388/5.99/42.2	cd01159 NcnH hydroxylase involved in the biosynthesis of polyketides. $1.26e^{-125}$	Monooxygenase *R. rhodochrous* WP_043798582.1 388 aa 352/388(91%)	Monooxygenase WP_011597918.1 388aa 285/388(73%)
KshA1	391/5.0/44.3	cd03531 Rieske_RO_Alpha_KSH $1.08e^{-82}$ COG4638 HcaE (ring-hydroxylating dioxygenase) $2.35e^{-35}$	3-ketosteroid-9-alpha-hydroxylase *R. rhodochrous* WP_059174956.1 391 aa 389/391(99%)	3-ketosteroid-9-alpha-hydroxylase oxygenase subunit WP_011597917.1 401 aa 304/392(78%)
HP	85/5.8/9.3	nd	Hypothetical protein *Rhodococcus sp.* P14 WP_010596083.1 85 aa 84/85(99%) Putative ferredoxin *R. rhodochrous* ADY18311.1 88 aa 50/85(59%)	WP_011596915 hypothetical protein 115 aa 39/112 (41.33%)
Short chain ox	263/5.9/28.0	cd05356 17beta-HSD1_like_SDR_c $4.00e^{-39}$ (non specific)	Short-chain dehydrogenase *R. rhodochrous* 263 aa 258/263(98%)	Short-chain dehydrogenase WP_011597916.1 263 aa 207/262(79%)

Table S1: b) ORFs near to KshA2

ORF	aa/ pI/ kDa	Protein domains/Evalue	Max amino acid ID	*R. jostii*
KstD2	564/7.1/6.0	PRK128453-ketosteroid-delta-1-dehydrogenase $0e^{+00}$ pfam00890 FAD binding domain $5.29e^{-67}$	3-ketosteroid-delta-1-dehydrogenase *R. rhodochrous* WP_054372667.1 569aa 551/563(98%)	3-ketosteroid-delta-1-dehydrogenase WP_011599743 578 aa 393/565(70%)
KshA2	388/4.8/43.6	cd03531 Rieske_RO_Alpha_KSH $8.05e^{-85}$ COG4638 HcaE $2.44e^{-33}$	3-ketosteroid-9-alpha-hydroxylase *R. rhodochrous* WP_054372666.1 388 aa 379/388(98%)	Probable dioxygenase Rieske iron-sulfur component ABG94295 397 aa 261/367(71%)
Hydroxylase	393/5.0/43.4	cd01159 NcnH (Naphthocyclinone hydroxylase) $3.34e^{-99}$	Hydroxylase *R. rhodochrous* WP_054372665.1 393aa 387/393(98%)	hydroxylase WP_011595215 393aa 260/394(66%)
Dioxygenase	301/4.9/32.8	cd07237 BphC1-RGP6_C_like (C-terminal domain of 2,3-dihydroxybiphenyl 1,2-dioxygenase BphC) $7.89e^{-71}$ cd07252 BphC1-RGP6_N_like (N-terminal domain of 2,3-dihydroxybiphenyl 1,2-dioxygenase BphC) $1.39e^{-41}$	2,3-dihydroxybiphenyl 1,2-dioxygenase *R. rhodochrous* WP_054372664.1 302aa 270/286(94%)	Iron-dependent extradiol dioxygenase WP_011595214 301 aa 222/285(78%)
IclR	341/10.7/25.8	Smart00346 HTH_ICLR (helix_turn_helix isocitrate lyase regulation) $4.24e^{-18}$ COG1414 IclR $3.18e^{-41}$	IclR family transcriptional regulator *R. rhodochrous* WP_054372663.1 241aa 225/238(95%)	IclR family transcriptional regulator WP_011599742 254 aa 144/232(62%)

Table S1: c) ORFs near to KshA3

ORF	aa/ pI/ kDa	Protein domains/Evalue	Max amino acid ID	*R. jostii*
hsaA	411/5.8/45.0	cd01159 NcnH Hydroxylase involved in the biosynthesis of polyketides. $3.69e^{-157}$	Monooxygenase *R. rhodochrous* WP_054371646.1 391aa 384/391(98%)	3-HSA hydroxylase, oxygenase ABG96325 1/382 (88.74 %)
KshA3	380/5.02/43.2	cd03531 Rieske_RO_Alpha_KSH $8.79e^{-85}$ COG4638 HcaE (ring-hydroxylating dioxygenase) $3.19e^{-36}$	3-ketosteroid-9-alpha-hydroxylase R. *rhodochrous* WP_054371647.1 380aa 374/379(99%)	3-ketosteroid-9-alpha-hydroxylase oxygenase subunit ABG96324 409 aa 317/377(84%)
Orf1	112/6.3/12.3	nd	Hypothetical protein *R. rhodochrous* WP_054371648.1 112aa 108/112(96%) Putative ferredoxin *R. rhodochrous* ADY18319.1 125 aa 74/118(63%)	Hypothetical protein WP_011596915 115aa 70/115(61%)
Orf 2	68/10.8/7.9	nd	Hypothetical protein *R. rhodochrous* WP_054371650.1 67aa 66/67(99%)	nd
Orf 3	214/4.5/22.8	COG0262 Dihydrofolate reductase [Coenzyme transport and metabolism] $7.43e^{-15}$	Deaminase *R. zopfii* WP_072816204.1 214aa 207/214(97%)	nd

Table S1: d) ORFs near to KshB

ORF	aa/ pI/ kDa	Protein domains/Evalue	Max amino acid ID	*R. jostii*
Alcohol DH	423/5.7/46.4	TIGR04266 NDMA-dependent methanol dehydrogenase $0e^{00}$	NDMA-dependent methanol dehydrogenase *R.rhodochrous* WP_059383282.1 423aa 422/423(99%)	Probable 1,3-propanediol dehydrogenase ABG97834 400 aa 373/400(93%)
KshB	351/4.5/37.7	cd06214 PA_degradation_oxidoreductase_like $3.57e^{-111}$ cd00207 fer2 (2Fe-2S iron-sulfur cluster binding domain) $2.01e^{-21}$	3-ketosteroid-9-alpha-hydroxylase *R. rhodochrous* WP_059174388 351 aa 350/351(99%)	3-ketosteroid-9-alpha-hydroxylase reductase subunit WP_011598098 351 aa 244/351(70%)
MFS transporter	416/10.7/41.6	cd06174 MFS $1.46e^{-08}$	MFS transporter *Rhodococcus* sp. P14 WP_010591968.1 415aa 414/415(99%)	nd

nd: non-detected; Ids under 40% were not considered.

Table S2: PredictProtein prediction of the secondary structure and solvent accessibility in % of the three KshAs from *R. ruber*

Secondary Structure %				Solvent Accessibility %			
Structure	**KshA1**	**KshA2**	**KshA3**	**Accessibility**	**KshA1**	**KshA2**	**KshA3**
Strand	17.65	17.78	18.95	Exposed	39.90	39.95	39.47
Loop	65.47	65.46	59.21	Buried	49.36	49.48	52.37
Helix	10.74	16.88	21.84	Intermediate	10.74	10.57	39.47

Figure S3. Growth of *R. ruber* strains in different substrates at 48h

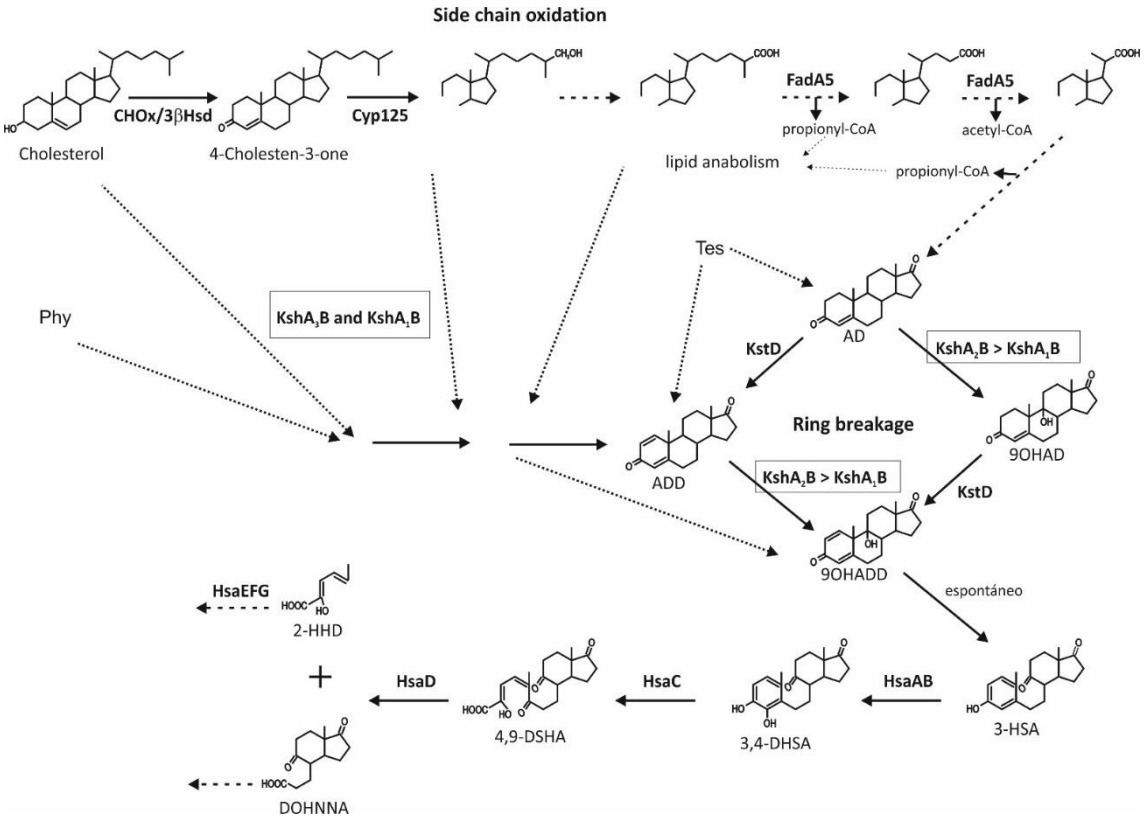


Figure S4. Scheme of aerobic cholesterol degradation in actinobacteria (Ouellet, Guan *et al.* 2010; García, Uhía *et al.* 2012; Yao, Xu *et al.* 2014; Bergstrand, Cardenas *et al.* 2016). Cholesterol pathway begins with its oxidation to cholestenone by the action of a cholesterol oxidase (ChoX) or a 3-β-hydroxysteroid dehydrogenase (HSD). The degradation continues with side chain cleavage via a mechanism similar to β-oxidation of fatty acids: in four cycles of β-oxidation, 3 mol of propionyl coenzyme A (propionyl-CoA) and 1 mol of acetyl-CoA are formed from the side chain. The steroid ring structure is degraded by oxygen-dependent opening and subsequent hydrolytic cleavage of rings A and B following the 9,10-seco pathway. In this central catabolic pathway AD is transformed into 9OH-ADD with the involvement of the network of KstD and KshAB activities. Intermediaries from upper pathway can be incorporated to the central pathway following alternative steps until yielding ADD or a derivative. After that, C and D rings are cleavage. The lower metabolic steps lead to central metabolites. The proposed pathway of testosterone degradation in *Comamonas testosteroni* TA441 (Horinouchi, Hayashi *et al.* 2003) or *Endozoicomonas montiporae* (Ding, Shiu *et al.* 2016) is via AD/ADD pathway. Similarly, DHEA catabolism in *C. testosteroni* follows the same pathway (Horinouchi, Kurita *et al.* 2010).The roles of *R. ruber* KshAs are also indicated in this scheme.

CAPÍTULO 4

Biotechnological applications of the genus *Rhodococcus*: generation of a dehydrogenation-enhanced strain

1. Background

Microorganisms have been used for C1 - C2 dehydrogenation of different 3- ketosteroids in order to obtain valuable intermediates in the synthesis of pharmaceutical compounds (Donova 2007). For instance, *Arthrobacter simplex* is commonly used industrially for the dehydrogenation of pregnanes, androstanes, and steroidal derivatives for more than 50 years (Nobile, Charney *et al.*, 1955).

Effective biotechnological methods are currently under development to obtain the dehydrogenation of compounds with yields over 90% (see Introduction section 6.1.5). It has been shown that the selectivity of dehydrogenation can be enhanced by overexpression of dehydrogenase activity in the appropriate host microorganism (Spelling 2008). In addition, it is noteworthy that although a wide variety of KstDs isoforms have been characterized from different sources, many of them are still unknown in both the specificity of the different dehydrogenation activities and their efficacy in the process.

The heterologous expression of KstDs in a host organism in order to obtain efficient production of the products of this enzyme has been successfully tested. For example in *Corynebacterium creatum* the KstD of *M. neoaurum* was overexpressed for the efficient transformation of AD in ADD (Zhang, Wu *et al.*, 2016) and in *Bacillus subtilis* the heterologous expression of a KstD of *Arthrobacter simplex* allowed to improve the activity of the enzyme 30 times more than in their original organism (Li, Lu *et al.*, 2007).

In order to obtain a strain of *Rhodococcus* with an enhanced dehydrogenation capacity and to study its biotechnological possibilities, we have proceeded to the overexpression of *kstD*s from different origins. *Rhodococcus erythropolis* CECT3014 has been used as an overexpression organism, due to its ability to electroporate and the existence of adequate expression vectors.

In this study, two *kstDs* genes were cloned in an expression vector to be over expressed in a *Rhodococcus* strain: *MSMEG_5941* encoding the KstD1 $_{MSMEG}$ protein of *Mycobacterium smegmatis* mc2 155 (Galán, Uhía *et al.*, 2016) and *stdH* from *Pseudomonas putida* (delta 4, 5α-steroid dehydrogenase). Our results showed that only KstD1 $_{MSMEG}$ could be overexpressed in *Rhodococcus erythropolis* 3014.

2. Results and Discussion

2.1. Expression of recombinant proteins

First, KstDs from different sources were chosen in order to clone them and get the dehydrogenation-enhanced *Rhodococcus* strain. The first option was to clone the KstDs from groups of collaboration in the steroid project: *MSMEG_5941* and *stdH.* These *kstD* genes were kindly donated by Jose Luis García (Environmental Biotechnology Lab, CIB) and Chema Luengo (Molecular Biology Department, University of León) respectively.

On the other hand, we decided to use the *Rhodococcus* strain CECT3014. It is able to be transformed with pTIP-QC1 as we cannot actually use any expression vector for *R. ruber.* Moreover, our laboratory has already use this strain for steroids research (Fernandez de las Heras, Mascaraque *et al.* 2011).

Therefore, for this study the two KstDs, one from *M. smegmatis* (*MSMEG_5941*) and another from *P. putida* (*stdH*) were cloned into the *Rhodococcus* expression vector pTIP-QC1. Its protein sequences are collected in supplemental material S1. The constructs obtained (Figure 1) were used to transform the electrocompetent cells of *R. erythropolis* CECT3014.

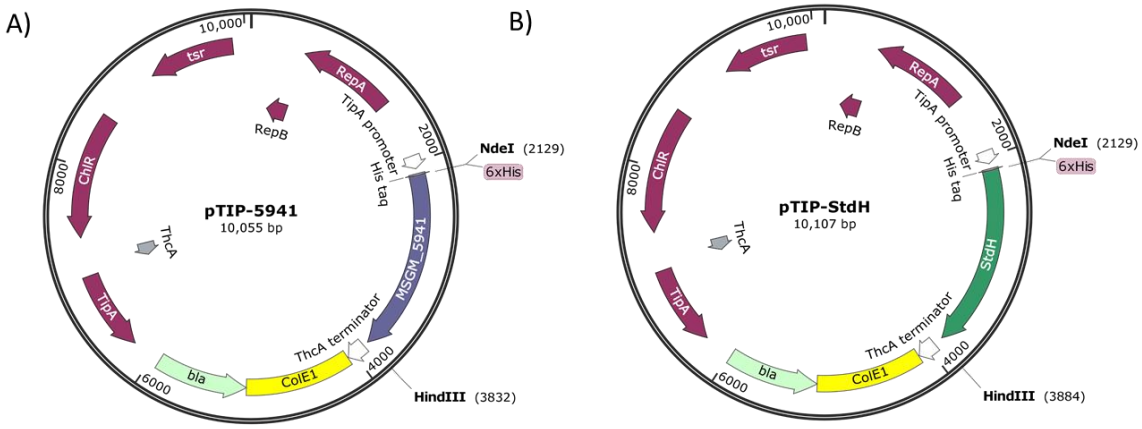


Figure 1: Scheme of pTIp vectors built in this work. The genes (*MSMEG_5941 and stdH*) were cloned into pTIP-QC1 in the *NdeI-HindIII* sites. A) pTIP-QC1 harboring *kstD* (*MSMEG_5941*) from *Mycobacterium smegmatis*; B) pTIP-QC1 harboring *kstD* (*stdH*) from *Pseudomonas sp.*

Transformation with the pTIP-StdH vector resulted in a massive growth despite repeated attempts of transformation. This result may be due to a recombination process with the *stdH* gene, which allows the incorporation of the whole plasmid into the genome and in that way

acquiring the cells the resistance to chloramphenicol. As discussed in the Introduction and Results (Chapter 1), *Rhodococcus* genomes are characterized by its ability of illegitimate recombination.

On the other hand, electroporation and transformation process was successful with the pTIP-5941 vector containing the *MSMEG_5941* gene *(*protein KstD1$_{MSMEG}$) Cell extracts from three colonies of *R. erythropolis* 3014 transformed harboring the plasmid pTIP-5941 were induced with thiostrepton and afterwards they were analyzed by SDS-PAGE (Figure 2). In the gel the high level of KstD1$_{MSMEG}$ protein was detected in the three colonies analyzed. These extracts were used for enzymatic activity assays.

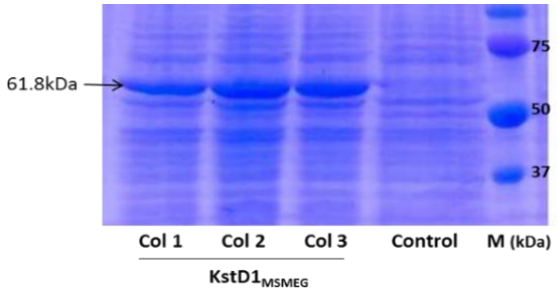


Figure 2. SDS PAGE (12.5%) analysis after staining with Coomassie Blue. Cellular extracts (20 µg of total protein) from *R. erythropolis* 3014 transformed with the vector pTIP-5941. Lanes 1 to 3: extracts from three different colonies (col) expressing *MSMEG_5941* (61.8 kDa protein); Lane 4: control extract (pTIP-QC1 empty); Lane 5: marker (M) Precision Plus Protein Standars (BioRad).

2.2. KstD1$_{MSMEG}$ activity

Once protein KstD1$_{MSMEG}$ was successfully obtained, its activity profiles in the presence of different substrates were analysed (Figure 3). As this strain keeps intact all its KstDs, we determined first the activity in an extract of *R. erythropolis* cells harboring the empty vector pTIP- QC1; it did not give detectable activity with any substrate (around 0.000001 U / mg). However when an extract of *R. erythropolis* overexpressing KstD1$_{MSMEG}$ was used, we could detect a higher activity with different substrates: 5α-testosterone followed by progesterone and testosterone. The activity obtained was hardly detectable in the presence of AD and DOC. KstD1$_{MSMEG}$ did not displayed detectable activity on: androsterone, 19OH-AD, 9OH-AD, corticosterone and DHEA.

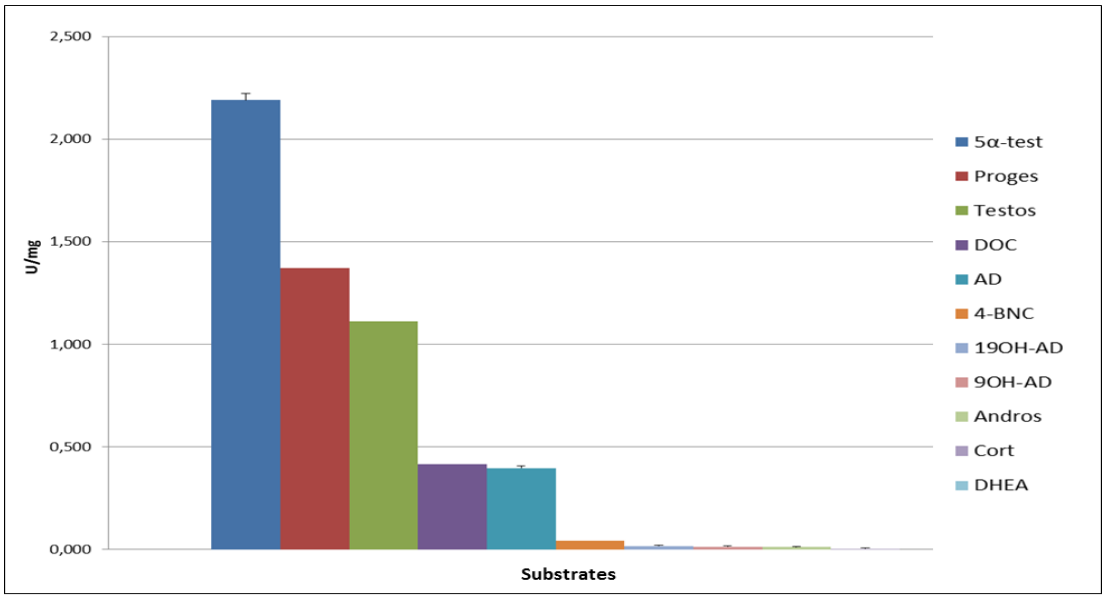


Figure 3. KstD activity in U/mg in *R. erythropolis* CECT3014 expressing KstD1$_{MSMEG}$. The activity in U/mg of KstD1$_{MSMEG}$ (by triplicate) for different substrates, using DCPIP as an artificial electron acceptor, was determined as indicated in Materials and Methods, section 4.3. Abbreviations: **5α-test**: 5α-testosterone; **Proges**: progesterone; Testosterone: testosterone; **DOC**: deoxycorticosterone; **AD**: 4-androsten-3,17-dione; **4-BNC**: 4-pregnen-3-one-20β-carboxylic acid; **19OH-AD**: 19-hydroxy-4-androsten-3,17-dione; **9OH-AD**: 9α-hydroxy-4-androsten-3,17-dione; **Andros**: androesterone; **Cort**: corticosterone; **DHEA**: dehiroepiandrosterone.

To confirm this data and for comparative purposes with other KstDs, kinetics studies were performed. These analyses were made for the substrates that showed activity in the preliminary approximation. The data obtained are in Table 1.

TABLE 1: Substrate profiles of *M. smegmatis* KstD1 expressed an analyzed in cell-free extracts of *R. erythropolis.* RCE: relative catalytic efficiency given by the ratio Activity/Km

Sustrates	U/mL	Km (μM)	RCE: U/(mL*μM)
5α-testosterone	7.0±0.6	122.9±17.5	0.05318687
Progesterone	0.92±0.05	350.1±72.3	0.00264328
Testostosterone	0.59±0.01	321.2±85.1	0.00186095
DOC	0.41±0.15	269.8±80.2	0.00152416
AD	0.35±0.01	49.6±10.8	0.00711525

KstD1$_{MSMEG}$ displays a 68% identity with KstD3, a 43% with KstD2 and a 35% with KstD1 from *R. ruber* Chol-4. In a similar way to KstD3 (in Chapter 2 of this work the preference of the *R. ruber* KstD3 by saturated steroidal substrates is shown.), the best substrates for KstD1$_{MSMEG}$ are saturated substrates such as 5α-testosterone. In *R. ruber* Chol-4 the enzymes KstD3 and KstD2 are related to cholesterol metabolism, since the double mutant of these enzymes is unable to

grow in cholesterol (Fernández de las Heras, van der Geize *et al.*, 2012). In *M. smegmatis* $KstD1_{MSMEG}$ is, also, the major enzyme involved in cholesterol catabolism (Uhía, Galán *et al.* 2012, Galán, Uhía *et al.* 2016).

In order to get a battery of functional KstDs for rhodococcal biotechnological applications, overexpression of other KstDs available from different sources (e.g. *MSMEG_2869* and *stdI*) and their subsequent characterization for comparative purposes are still pending.

Supplementary material.

S1. Protein sequences of the KstDs cloned:

A) KstD1: source *M. smegmatis* mc^2 155 (MSMEG_5941).

MTGQEYDVVVVGSGAAGMVAALTAAHQGLSTVVVEKAPHYGGSTARSGGGVWIPNNEILKRDGVKDTP
DEARKYLHAIIGDVVPAEKIDTYLDRGPEMLSFVLKHSPLKLCWVPGYSDYYPETPGGKPTGRSVEPKPFDAN
KLGPDLKGLEPPYGKVPMNMVVMQQDYVRLNQLKRHPRGVLRSLKVGIRATWGKVSGKNLVGMGRALI
APLRIGLREAGVPVLLNTALTDLYVEDGRVLGIYVRDTTAGDDAEPRLIRARHGVILGSGGFEHNEQMRVKY
QRAPITTEWTVGAVANTGDGIVAAEKLGAALELMEDAWWGPTVPLVGAPWFALSERNSPGSIIVNMSGK
RFMNESMPYVEACHHMYGGQYGQGPGPGENIPAWLIFDQQYRDRYIFAGLQPGQRIPSKWLESGVIVKA
DTLVELAEKTGLPADQFTSTIERFNGFARSGVDEDFHRGESAYDRYYGDPTNKPNPNLGEIKHGPFYAAKMV
PGDLGTKGGIRTDVRGRALRDDDSVIEGLYAAGNVSSPVMGHTYPGPGGTIGPAMTFGYLAALDIAATAAA
SRKG

B) StdH: source *Pseudomonas putida*

MTAQVQSQSAYDVIVVGSGAGAMTSAVFLADHGLRVLIVEKSDKYGGTSAISGGGIWIPNNHYFARMGG
NDSPELARRYLQASAGEHVEPARLEAYLANAPKMIEALTRNSRVRYAVAAKYPDYYPHLPGSLAGGRTLDPE
LFDTSLLGEELNNLRKPSPSTLLMGRIAWIARHAHKAMARNFGWRLLIMGLMLRYKLDFKWRKKSKYDRRA
ALGSSLVASLRRSLMDRNVPLWLNTDFRELLMKDGKVAGIVVERDGQEQRIEARHGVIFGSGGFEQNQAL
REQYLPKPTQMAWSATPPGNNTGTALLAGMKEGAATALMDWAWWAPTIAVPGEDKPRGIFAERAFPG
AIVVNSLGRRFVNEAAPYLEFVDAMHRDNQKTGGCSVPAWVIFDAHFRFHYAMGPLMPAQVMPDSRLR
KEWLNTLYWKADTLAGLAAQIGVDADGLENTVVKVNDYARSGVDPDFGRGGNVFDRYYGDCNIKPNPCL
APLRKGPFYAMRLDAGDIGTKGGLLTNEHAQVVREDGTPIAGLYAIGNCSASVMGTSYPGAGGTLGPAMT
FGYVAANHLAREAS

VII. DISCUSIÓN INTEGRADORA Y APORTACIONES FUNDAMENTALES DE ESTA TESIS

DISCUSIÓN INTEGRADORA Y APORTACIONES FUNDAMENTALES DE ESTA TESIS

1. Genoma de *Rhodococcus ruber* Chol-4

En trabajos previos realizados en el laboratorio de Ingeniería Genética y Metabólica del departamento de Bioquímica y Biología Molecular I de la UCM se aisló e identificó la cepa *Rhodococcus rube*r Chol-4. Esta cepa se obtuvo a partir de una muestra de lodos proveniente de una planta de tratamiento de aguas residuales (Ciudad Real). *R. ruber* mostró tener capacidades metabólicas para la degradación de compuestos esteroideos (Fernandez de las Heras, García Fernández *et al.* 2009), por lo que esta cepa presenta interés biotecnológico no solo para posibles aplicaciones en biorremediación sino también para la producción de sintones o compuestos bioactivos para la industria farmacéutica. Con el objetivo de caracterizar y determinar los genes y las rutas involucradas en la degradación de esteroides en esta cepa se hizo una primera secuenciación en la que se identificaron por homología varios genes relacionados con el metabolismo de esteroides tales como las colesterol oxidasas y las KstDs (Fernandez de las Heras, Mascaraque *et al.* 2011, Fernández de las Heras, van der Geize *et al.* 2012, Fernández de las Heras, Perera *et al.* 2014).

En el presente estudio se ha realizado una segunda secuenciación del genoma de *R. ruber* para mejorar la resolución de la primera secuenciación y obtener el menor número de *contigs* posibles. El ensamblaje final dio como resultado 50 *contigs* que cubren 5.4 Mb con un 70.7% de G+C. La notación genómica hecha con el servidor RAST identificó 5049 secuencias codificantes y 60 secuencias de RNA. El genoma de la cepa Chol-4 es de tamaño medio (5.4 Mb) sin embargo muestra una redundancia de genes y diversidad de rutas metabólicas lo que puede contribuir a que *R. ruber* sea capaz de utilizar simultáneamente compuestos que se encuentran en su entorno como se ha visto en otras especies del género (van der Geize y Dijkhuizen 2004, Fernández de las Heras, van der Geize *et al.* 2012, Dijkstra, van Oosterwijk *et al.* 2016).

Una característica notable del genoma de varias especies de *Rhodococcus* es la presencia de plásmidos, tanto circulares como lineales o una mezcla de los dos. Los plásmidos identificados hasta la fecha se han relacionado con funciones de transposición que dotan de movilidad al genoma (Denis-Larose, Labbe *et al.* 1997, Kalkus, Menne *et al.* 1998, Warren, Hsiao *et al.* 2004, Yang, Lessard *et al.* 2007, Yang, Lessard *et al.* 2007), además de contener genes que enriquecen las capacidades metabólicas de las cepas en las que se encuentran (Denis-Larose, Labbe *et al.* 1997, O'Brien, Parker *et al.* 2002, König, Eulberg *et al.* 2004, Priefert, O'Brien *et al.* 2004). Sin embargo, no fue posible determinar la presencia de ningún tipo de plásmido en *R. ruber*, aunque

una característica notable fue la identificación de genes relacionados con promover la alta frecuencia de recombinaciones ilegítimas, además de la presencia de varios transposones (Capítulo 1, Tabla suplementaria S2). Esta característica explica en cierta medida las dificultades técnicas que el trabajo con esta cepa presentó en lo relacionado a la transformación con plásmidos: *R. ruber* es una cepa que aunque conjuga muy bien, se electropora con dificultad y en ambos casos acaba recombinando la mayoría de plásmidos que incorpora. Por otra parte, a pesar de diversos intentos para obtener un plásmido de expresión en esta cepa, ninguno de ellos ha sido hasta ahora fructífero.

Gracias al análisis del genoma de *R. ruber* se ha identificado la presencia de genes que codifican rutas de degradación de compuestos aromáticos tanto centrales como periféricas. Las rutas periféricas se encargan de descomponer compuestos de degradación compleja (p. ej. colesterol o benzoato) en intermediarios comunes (p. ej. catecol o ácido gentísico) que a su vez son degradados mediante las vías centrales de degradación de compuestos aromáticos para transformarlos en intermediarios que entran en las rutas básicas como el ciclo de Krebs.

En *R. jostii* RHA1 se han identificado ocho *clusters* génicos de rutas centrales de degradación de compuestos aromáticos (Yam, van der Geize *et al.* 2010), de las que siete se encuentran en el genoma de *R. ruber*. Estos siete *clusters* son: del β-cetoadipato, 2-hidroxipentadienoato, gentisato, homogentisato (Hmg), hidroxiquinol, homoprotocatecuato y un *cluster* cuyos sustratos aún no se han identificado. Además, se han detectado varias rutas periféricas relacionadas con la degradación del: benzoato, isopropilbenzeno, esteroides, vanilato, naftaleno, acetofenona y aminoácidos. Los agrupamientos de genes que codifican para estas rutas son sintónicos puesto que la distribución y orientación espacial entre ellos es muy similar a la encontrada en *R. jostii* y otras especies del género.

R. ruber fue capaz de crecer en presencia de varios de los compuestos cuyos *clusters* génicos de catabolismo fueron identificados en su genoma, por ejemplo benzoato, naftaleno o gentisato. Adicionalmente la generación de dos mutantes en las rutas de degradación del naftaleno (genes *nar*) y del protocatecuato (genes *pca*) nos permitió comprobar que dos de los *cluster* previamente identificados *in silico* se correspondían con la ruta de degradación propuesta.

2. Enzimas involucradas en la degradación del anillo esteroideo A

De todas las rutas identificadas, la ruta de degradación de esteroides tiene especial interés debido a que los intermediarios de esta vía son compuestos muy apreciados en la industria farmaceútica. Con el fin de focalizar nuestro trabajo en puntos estratégicos de la ruta de

degradación de los esteroides, se continuó con el estudio y caracterización de las enzimas involucradas en el proceso de ruptura de los anillos A y B, en concreto de la 3-cetoesteroide-Δ^1-deshidrogenasa (KstD) y la 3-cetoesteroide-9α-hidroxila (KshAB) (Itagaki, Hatta *et al.* 1990, van der Geize, Hessels *et al.* 2000, Knol, Bodewits *et al.* 2008, Hu, Van Der Geize *et al.* 2010, Capyk, Casabon *et al.* 2011, Donova and Egorova 2012, Dijkstra, van Oosterwijk *et al.* 2016). Las KstDs se encargan de la transformación de 4-eno-3-oxo-esteroides (p. ej. AD) a 1,4-dieno-3-oxo-esteroides (p. ej. ADD) (Choi, Molnar *et al.* 1995, van der Geize, Hessels *et al.* 2000, Brzostek, Sliwinski *et al.* 2005, Fernández de las Heras, van der Geize *et al.* 2012), mientras que las KshAB catalizan la hidroxilación de 4-eno-3-oxo-esteroides o 1,4-dieno-3-oxoesteroides (p. ej. AD y ADD) en 9-hidroxi-4-eno-3-oxo-esteroides o 9-hidroxi-1,4-dieno-3-oxo-esteroides (p. ej. 9OH-AD o 9OH-ADD) (van der Geize, Hessels *et al.* 2002, Petrusma, Dijkhuizen *et al.* 2009, Capyk, Strynadka *et al.* 2010, Hu, Van Der Geize *et al.* 2010, Petrusma, van der Geize *et al.* 2014).

Previamente en nuestro grupo de trabajo se había identificado e iniciado la caracterización de tres KstDs en *R. ruber* Chol-4 (Fernández de las Heras, van der Geize *et al.* 2012). En esta Tesis Doctoral se ha continuado el estudio de estas enzimas y además se ha llevado a cabo la identificación y caracterización funcional de tres KshAs y una KshB en *R. ruber*. La presencia de múltiples isoformas de un mismo gen en *R. ruber* es consistente con la gran versatilidad metabólica descrita en las especies del género *Rhodococcus* que también cuentan con un número variable de isoenzimas (van der Geize, Dijkhuizen, 2004; Yam, Okamoto, 2011). Por ejemplo en *R. jostii RHA1* se han detectado al menos seis KstDs y cuatro KhAs (Mathieu, Schloendorn *et al.* 2008), en *R. erythropolis* SQ1 se han encontrado tres genes *kstD*s y tres *kshA*s (van der Geize, Hessels *et al.* 2008, van der Geize, Hessels *et al.* 2001, Knol 2011), en *R. rhodochrous* se han caracterizado cinco KshAs (Wilbrink, Petrusma *et al.* 2011) y tres KstDs (Liu, Shen *et al.* 2016, Morii, Fujii *et al.* 1998).

Tanto KstD como KshAB requieren de la unión de coenzimas redox, FAD y NAD respectivamente. Las KstDs tienen una sola unidad en la que se encuentran tanto el sitio de unión al sustrato como el de unión a la coenzima. Por otro lado, KshAB se divide en dos unidades: la unidad KshA que alberga el centro activo y la unidad KshB de unión al NAD, cada una codificada por un gen distinto. La interacción que existe entre estas dos unidades (KshA y KshB) parece ser poco específica, ya que en *R. ruber* hemos podido identificar solo un gen *kshB* frente a tres genes *kshAs,* lo que es común en genomas donde se han identificado y estudiado estas enzimas. Por ejemplo en los genomas de *R. rhodochrous* DSM43269 (Petrusma, Dijkhuizen *et al.* 2009), *R. jostii* RHA1 (Mathieu, Schloendorn *et al.* 2008), *R. erythropolis* SQ1 (van der Geize, Hessels *et al.*

2002) o *G. neofelifaecis* NRRL B-3805 (Zhang, Ren *et al.* 2015), existe un mayor número de KshAs que de KshBs. Es más, se ha conseguido en mutantes de *Mycobacterium sp.* NRRL B-3805 que la actividad KshAB sea restituida mediante la expresión heteróloga del gen *kshA* de *M. smegmatis* mc^2155 y el gen *kshB* de *G. neofelifaecis* NRRL B-3805 (Yuan, Chen *et al.* 2015).

Un paso inicial en la caracterización de las enzimas KstD y KshA fue el modelado molecular mediante el programa I-TASSER. El modelado se basó en la estructura cristalográfica de KstD1 de *R. erythropolis* SQ1 (Rohman, van Oosterwijk *et al.* 2013) y de KshA5 de *R. rhodochrous* DSM43269 (Penfield, Worrall *et al.* 2014) que eran las más parecidas a las isoenzimas de *R. ruber*. Una vez obtenido el modelo tridimensional de cada proteína se procedió al alineamiento entre isoenzimas. Los resultados de esta Tesis Doctoral sugieren que los residuos que conforman el dominio catalítico en las tres KstDs y las tres KshAs de *R. ruber*, aunque altamente conservados, varían en cuanto a orientación y posición, lo que podría explicar en cierta medida su diferencia en términos de actividad o funcionalidad frente a diferentes sustratos (Figura 6 del Capítulo 2 y Figura 3 del Capítulo 3).

3. Contexto génico de las enzimas KstDs y Kshs en *R. ruber*

Cada gen de *kstDs* o *kshAB* se encuentra en un contexto génico diferente, los *clusters* en los que se halla cada gen pueden estar relacionados con el papel que desempeñan en la degradación de un determinado compuesto. Cabe destacar que los genes *kshA2* y *kstD2* se encuentran en un mismo *cluster* génico de la ruta de degradación de esteroides (Figura 1 del Capítulo 3) y son precisamente estas dos enzimas las que en los resultados presentados en esta Tesis Doctoral ha mostrado desempeñar un papel importante en la degradación del AD (Capítulo 2 y 3). Por otra parte, el gen de la *kshA3* se encuentra en el mismo *cluster* que los genes que codifican para las enzimas HsaC, HsaD y HsaA. Estas enzimas están involucradas en la ruta de degradación de los esteroides una vez han actuado las KstDs y KshABs (Horinouchi, Hayashi *et al.* 2005, van der Geize, Yam *et al.* 2007, Yam, D'Angelo *et al.* 2009). El gen *kstD1* se encuentra en el mismo *cluster* de genes que podrían estar relacionados con la degradación de la cadena lateral, tales como un *cyp450*, *fad5* y genes que codifican acil CoA hidratasa o transferasa. Mientras que el *cluster* en el que se encuentra *kstD3*, en cambio, es en el que se encuentran los genes *hsaF*, *hsaG* y *hsaE* relacionados con la ruta baja de degradación de esteroides (Kieslich 1985, Horinouchi, Hayashi *et al.* 2003, van der Geize, Yam *et al.* 2007). En contraste, el *cluster* de *kshB* no presenta genes que se hayan identificado como relacionados con la degradación de esteroides. La interacción de KshB con las kshAs no es específica por tanto no es de extrañar que no se encuentre en un *cluster* específico.

4. Papel fisiológico de los homólogos de *kshA* y *kstD* en el catabolismo de diferentes clases de esteroides.

A continuación, dada la redundancia genética en las enzimas de degradación de esteroides comentada anteriormente, se procedió a confirmar la importancia biológica y/o la funcionalidad de cada isoenzima. Además de contar con los mutantes de *kstD* previamente construidos en el laboratorio (Fernández de las Heras, van der Geize *et al.* 2012), se generaron mutantes de deleción génica de *kshAs* o *kshB.* Los resultados obtenidos con las cepas mutantes en los ensayos de crecimiento en presencia de diferentes esteroides como única fuente de carbono demostraron que la sola mutación del gen *kshB* es suficiente para impedir el crecimiento relacionado con la degradación de los anillos de esteroides, pero no detiene el crecimiento por la degradación de la cadena lateral pues se puede observar un crecimiento incipiente en sustratos como el colesterol o los fitosteroles. Este resultado es opuesto al obtenido en *R. erythropolis* SQ1 *(van der Geize, Hessels et al. 2002)* en el que, para que la degradación de la cadena lateral ocurra era necesaria la actividad de KshB, por tanto *R. ruber* parece tener un sistema de degradación de la cadena lateral independiente al de la ruptura de los anillos A y B.

Por otra parte, los mutantes de *R. ruber* que carecen de actividad KstD2 y KstD3, en la que KstD1 sería la única isoenzima funcional (Fernández de las Heras, van der Geize *et al.* 2012) o KshA1 y KshA3, en la que KshA2 sería la única isoenzima funcional, crecen muy poco y con dificultad en colesterol o fitoesteroles siendo el crecimiento similar al que se aprecia en el mutante en *kshB* o en la triple mutante KshA1A2A3. Estos resultados llevan a pensar que las isoenzimas KstD2, KstD3, KshA1 y KshA3 estarían involucradas en la degradación de esteroles con cadena lateral larga.

En el caso del AD, un esteroide sin cadena lateral, la ausencia de la actividad KstD2 es suficiente para detener el crecimiento, mientras que es necesaria la mutación de los dos genes *kshA1* y *kshA2* para obtener el mismo efecto en *R. ruber.* En coherencia con estos resultados, la sola presencia de las isoformas KshA3 o de KstD1 y KstD3, no es suficiente para que se dé el crecimiento de esta cepa en presencia de AD.

Hay que destacar que el crecimiento cuando KshA1 es la única enzima KshA funcional se ve ralentizado con respecto al crecimiento de la cepa silvestre, lo que podría deberse a que se requiera de un periodo de tiempo para la expresión adecuada de *kshA1* como los resultados de patrones de expresión, que se discuten más adelante, sugieren.

En cuanto al crecimiento en testosterona, otro esteroide de cadena lateral corta, es necesaria la presencia de KshA2 y KstD2 para que *R. ruber* pueda crecer. Sin embargo, se aprecia algo de crecimiento en testosterona cuando la KstD2 no está presente en el genoma similar incluso al crecimiento en el triple mutante de las KstDs. Este crecimiento residual sugiere que para la degradación de este esteroide pueden estar implicadas otro tipo de enzimas aunque no tan eficientes como la KstD2. Las isoenzimas KstD2 y KshA2 estarían por tanto involucradas en la degradación de esteroles de cadena lateral corta.

Por otra parte, en la degradación del ácido cólico los mutantes de *R. ruber* en *kshA1* o *kstD2* fueron incapaces de crecer en presencia de este sustrato como fuente de carbono hasta las 48 horas. Sin embargo, a partir de las 72 horas los mutantes en *kstD2* e incluso el triple mutante *ΔkshA1,2,3* recuperan la capacidad de crecer. Este hecho podría deberse a que se podría abrir una ruta alternativa de degradación en la que intervengan otras deshidrogenasas u otras actividades enzimáticas.

El caso de la kstD1 debe ser estudiado con más detenimiento puesto que en ningún resultado de esta Tesis Doctoral o en trabajos previos hechos en *R. ruber* se ha podido identificar funcionalidad *in vivo* con la degradación de esteroides de cadena larga o corta, a pesar de que su sustrato óptimo sea el 9OH-AD y sea capaz de utilizar el AD en ensayos *in-vitro* (Capítulo 2, Tabla 1). Sin embargo, los resultados de los crecimientos muestran que sólo puede ejercer su acción cuando está KstD2 presente en la célula. Posteriormente se discutirá este punto con más detalle.

En la Figura 1 se recoge un resumen esquemático de las aportaciones fundamentales de esta tesis en este apartado.

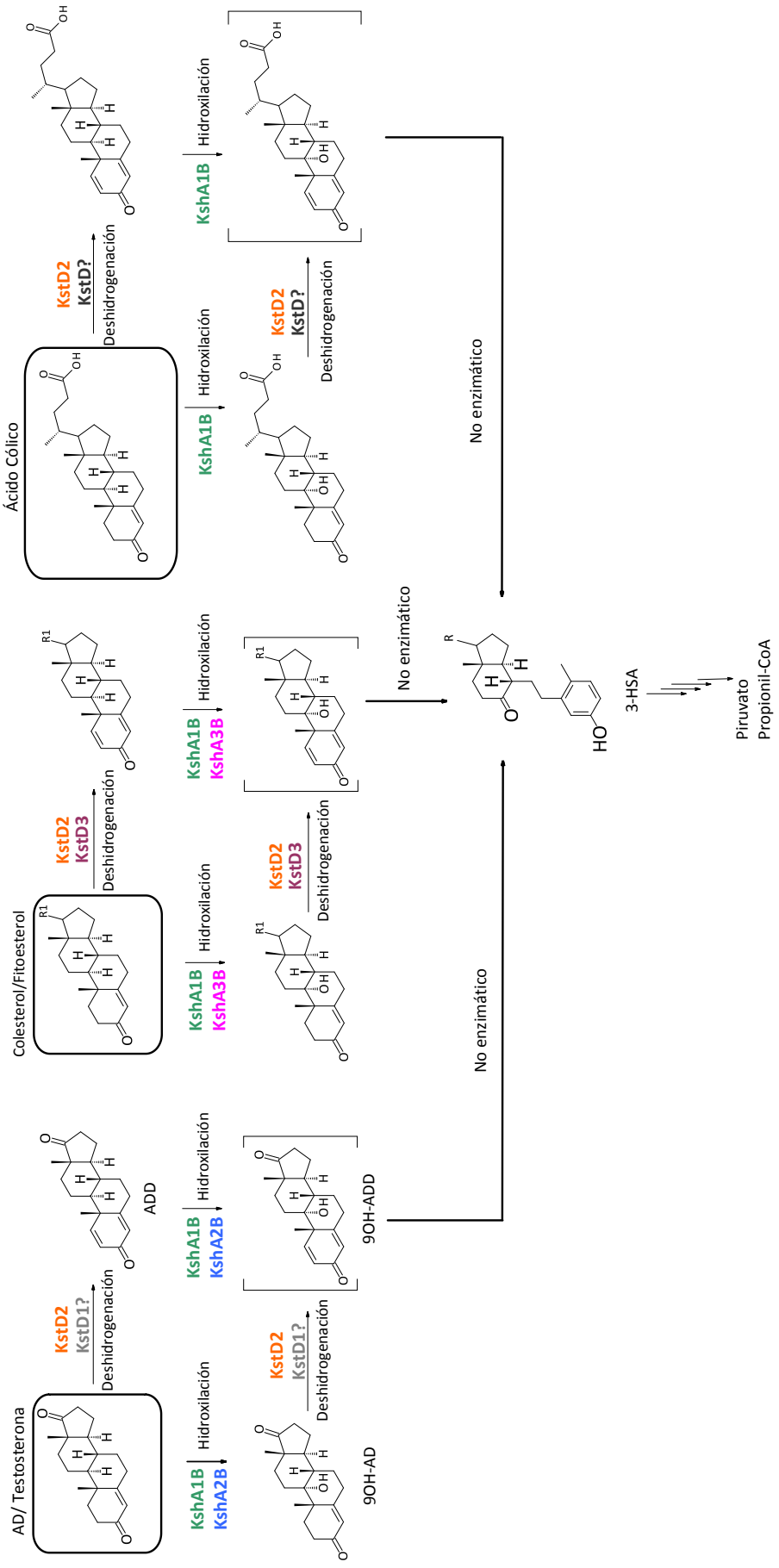


Figura 1. Resumen de la participación de las KstDs y KshAs de *R. ruber* Chol-4 en la ruta de degradación de esteroides de cadena larga, corta y ácido cólico. Varias flechas indican procesos multienzimáticos.

5. Estudios de expresión de las *kstDs* y *kshAs* en *R. ruber*.

Con objeto de profundizar en el estudio de la funcionalidad de las KstDs y KshAs en la degradación de esteroides en *R. ruber*, se llevó cabo el análisis de su expresión génica. Los resultados de esta Tesis Doctoral muestran que tanto los genes *kstDs* como los *kshAs* presentan una expresión constitutiva ya que se expresan en todas las condiciones de crecimiento ensayadas (AD, colesterol y acetato de sodio). Frente a esta expresión basal hemos observado una alta inducción adicional en presencia de esteroides. En presencia de AD se observa inducción únicamente de *kshA2* y de *kstD1*, pero no parece haber inducción de *kstD2* (a parte de la basal), lo que es curioso pues KstD2 es indispensable para el crecimiento de *R. ruber* en AD. En otros organismos se han publicado diferentes perfiles de expresión génica, como es el caso de *R. rhodochrous,* en donde tanto *kstD2* como *kstD1* se encuentran altamente inducidos en AD (Liu, Shen *et al.* 2016). En este organismo la enzima homologa de KstD2 de *R. ruber* es la principal enzima involucrada en la degradación del AD; sin embargo la enzima homologa de KstD1 también participa en la degradación del AD, aunque en menor grado. Al igual que en *R. ruber* la KstD3 de *R. rhodochrous* parece no participar en la degradación del AD y no se induce (Liu, Shen *et al.* 2016).

En presencia de colesterol como única fuente de carbono se aprecia una inducción de la expresión de *kshA2*, *kshA3* y *kstD3*. Dada la participación de KstD3 y KshA2 en la degradación del colesterol no sorprende que estos dos genes se encuentres inducidos, sin embargo llama la atención la inducción de kshA2, cuya presencia no es necesaria para la degradación del colesterol.

El caso de KstD1 merece un estudio aparte. Como se comentó previamente, no se ha podido determinar en qué ruta catabólica interviene *KstD1*. Los estudios de expresión *in vitro* nos indican que el gen *kstD1* se induce en presencia de AD y algo menos en colesterol (ver Figura 5 del Capítulo 2).

Un caso similar se ha encontrado en *R. erythropolis* strain SQ1 pero con las actividades KshA. El gen *kshA2* se induce en AD pero la cepa mutante en KshA1 que mantiene la actividad de KshA2B es incapaz de crecer en AD (van der Geize, Hessels *et al.* 2008). En este caso en concreto, un cambio de promotor en el gen *kshA2* bastó para que KshA2 complementara al mutante en KshA1. En el caso de *R.ruber*, la expresión del mensajero de *kstD1* en el mutante *kstD2* no se ve anulada pero sí alterada (Figura 5 del capítulo 2), dando una inducción mayor por colesterol que por AD. El hecho de que se pierda la inducción de *kstD1* en AD, podría justificar la ausencia de

crecimiento detectada en el mutante en *kstD2* en presencia de este esteroide. KstD1 podría ser una enzima que actué como refuerzo de las otras KstDs y cuya función en el crecimiento en AD dependería de la presencia de KstD2. Una función auxiliar en las KstDs se ha detectado en *Mycobacterium smegmatis* en donde la sobreexpresión del gen *kstD2* (MSMEG_4870) –no relacionado con el metabolismo del colesterol- fue capaz de revertir la incapacidad para crecer en colesterol del mutante en KstD1 (ΔMSMEG_5941) (Brzostek, Sliwinski *et al.* 2005). Queda pendiente por tanto otro tipo de estudios que lleven a analizar la funcionalidad de KstD1 en *R. ruber* y la naturaleza de la relación entre KstD1 y KstD2 en *R. ruber*.

La expresión diferencial de cada uno de los genes *kstDs* y *kshAs* en diferentes sustratos demuestra que es un mecanismo finamente regulado y que depende de cada cepa.

6. Propuesta de una ruta nueva de degradación de esteroides en *R. ruber*.

El género *Rhodococcus* tiene la habilidad de crecer en presencia tanto de esteroides de cadena larga como de cadena corta (Watanabe, Shimizu *et al.* 1986, Aihara, Watanabe *et al.* 1988, Ahmad, Roy *et al.* 1993, van der Geize, Hessels *et al.* 2000, Navas, González-Zorn *et al.* 2001, De Carvalho and Da Fonseca 2005, van der Geize, Yam *et al.* 2007, Fernandez de las Heras, García Fernández *et al.* 2009, Yam and Okamoto 2011). Sin embargo en la mayoría de las publicaciones sobre la degradación de esteroides en *Rhodococcus* parten de compuestos ampliamente aceptados como intermediarios de la ruta de degradación del colesterol como el AD, ADD o 9OH-AD (van der Geize, Hessels *et al.* 2000, van der Geize, Hessels *et al.* 2001, van der Geize, Hessels *et al.* 2002, van der Geize, Hessels *et al.* 2002, Yang, Lessard *et al.* 2007, Knol, Bodewits *et al.* 2008, van der Geize, Hessels *et al.* 2008, Petrusma, Dijkhuizen *et al.* 2009, Fernández de las Heras, van der Geize *et al.* 2012, Liu, Shen *et al.* 2016). Hasta la fecha las pocas publicaciones que identifican intermediarios a partir de la degradación del colesterol en *Rhodococcus* lo hacen en relación con la degradación de la cadena lateral (Yeh, Kuo *et al.* 2014). A pesar del diseño de cepas de Rhodococcus para la acumulación de a AD o ADD a partir de fitoesteroles o esteroles de cadena larga (van der Geize, Hessels *et al.* 2002, van der Geize, Hessels *et al.* 2008), solo existe una publicación en la que se obtiene ADD, aunque la mayor parte de producto obtenido es el 1,4-BNC (Yeh, Kuo *et al.* 2014). Esto pone de manifiesto una vez más que los intermediarios de esteroides de cadena larga parecen no pasar mayoritariamente por el AD, ADD o 9OHAD en *Rhodococcus*.

Sin embargo en *Micobacterium* existen varias publicaciones sobre la degradación del colesterol (Chipley, Dreyfuss *et al.* 1975, Andor, Jekkel *et al.* 2006, Brzostek, Dziadek *et al.* 2007, Yang,

Nesbitt *et al.* 2009, Ouellet, Johnston *et al.* 2011, Ivashina, Nikolayeva *et al.* 2012, Uhía, Galán *et al.* 2012, Yang, Lu *et al.* 2015, Wrońska, Brzostek *et al.* 2016) pero muy pocas relacionadas con la capacidad de micobacterias para degradar compuestos como el AD o el ADD directamente (Brzostek, Sliwinski *et al.* 2005, Li, Ge *et al.* 2016) ya que es difícil lograr el crecimiento de Micobacterias en presencia de AD o 9OH-AD. Recientemente en un esfuerzo por esclarecer la ausencia de crecimiento en cepas de *Mycobacterium* en presencia de estos intermediarios de la ruta de degradación del colesterol, se ha llegado a la identificación de un *cluster* alternativo a la ruta clásica propuesta para la degradación del colesterol (Fernández-Cabezón, García-Fernández *et al.* 2017). Este *cluster* se ha identificado en *M. smegmatis* y otras especies no patógenas del género (*M. vanbaalenii* PYR-1 o *M. neoaurum* DSM 44074) y ha sido caracterizado como uno de los responsables de la expresión de genes relacionados con el metabolismo de compuestos C-19 (considerados como tales esteroides sin cadena lateral en C-17, e.g. AD, ADD o 9OHAD). Entre estos genes se encuentra PadR, un regulador transcripcional que reprime la expresión de los genes indispensables para la degradación de compuestos como el AD o el ADD.

Es de notar que en *R. ruber,* el *cluster* donde esta KshA2 y KstD2 (Figura 1 del Capítulo 3), las dos enzimas mayormente involucradas en la degradación de AD*,* se encuentra un regulador PadR (D092_RS19405) y un regulador IcIR al igual que el *cluster* identificado en *M. smegmatis*.

Los resultados presentados en esta Tesis Doctoral (e.g. ausencia de intermediarios tipo AD o ADD por TLC en los mutantes de Ksh y KstD creciendo en colesterol y el estudio de crecimiento de los mutantes en distintos esteroides) inciden en que la ruta vía AD no es esencial para la degradación del colesterol en *R. ruber*, y que debe existir otra ruta vía KstD3 y KshA3B alternativa. El estudio más detallado de esta ruta nos permitirá saber si el *cluster* C19 vía padR estaría o no involucrado en esta ruta en esta cepa.

7. Aplicaciones Biotecnológicas de *R. ruber* Chol-4

Los rhodococci son organismos particularmente interesantes para la producción de esteroides a través de los procesos de bioconversión y biotransformación. La producción de intermediarios metabólicos que pueden ser usados como sintones o compuestos bioactivos de bajo coste tiene un gran potencial comercial. En este sentido, las actinobacterias como *Rhodococcus* son buenas candidatas para la producción de sintones debido a que: i) las cepas de *Rhodococcus* no son patógenas en su mayoría, ii) son relativamente fáciles de cultivar, iii) cuentan con herramientas de manipulación genética, iv) poseen un conjunto completo de enzimas catabólicas y v) su investigación funcional aumenta constantemente. (Ahmad, Roy *et al.* 1993, Ivshina, Grishko *et*

al. 2004, Van Der Geize, Hessels *et al.* 2008). En la Figura 2, se muestran esquemas de los diferentes mutantes obtenidos en este estudio y en trabajos llevados a cabo en este grupo.

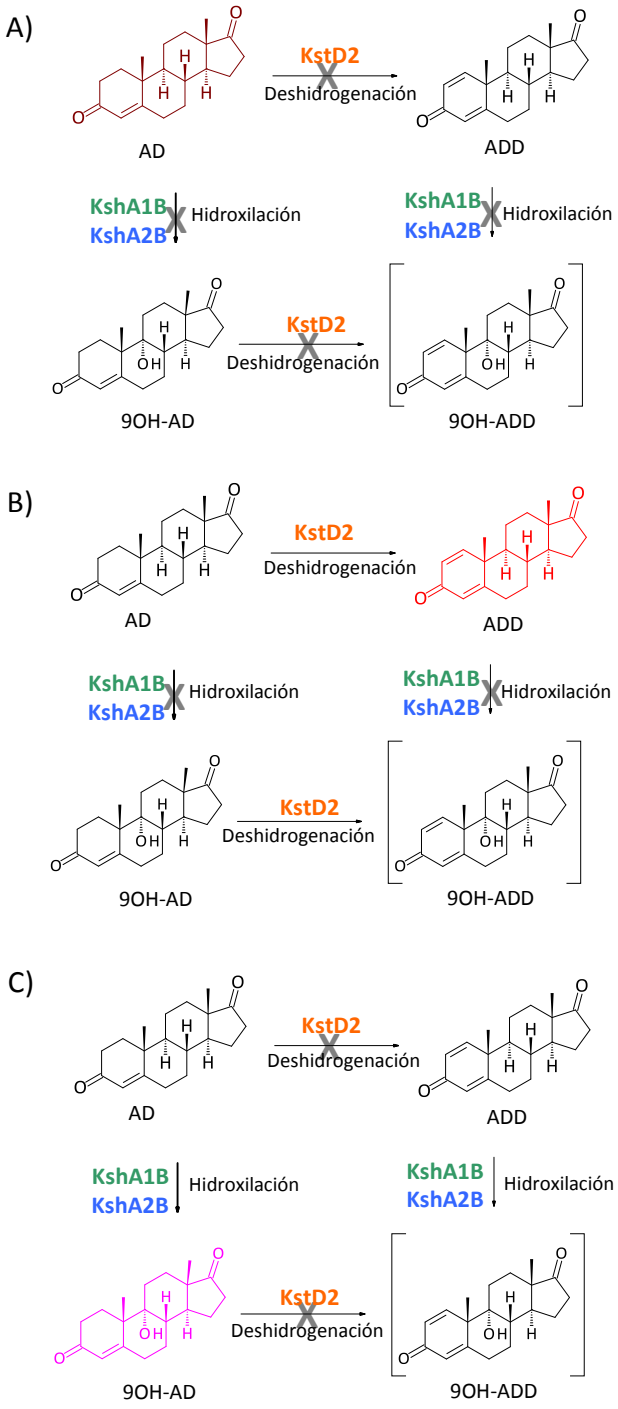


Figura 2. Esquema de las rutas interrumpidas por medio de mutagénesis dirigida en *R. ruber* Chol-4. Con cada ruta se pueden acumular un tipos de compuestos a partir de AD: a) Utilización hipotética de AD para su modificación por otras actividades enzimáticas que se aporten en *trans* en la propia bacteria usada como factoría; para ello se deben mutar las actividades propias de la cepa b) acumulación de ADD mediante la mutación de las actividades de Ksh involucradas en este paso y c) acumulación de 9OH-AD causada por la mutación de las actividades KstD responsables de este paso. Cada compuesto que se puede acumular aparece en color.

En resumen el conocimiento generado en la presente Tesis Doctoral y en trabajos previos llevados a cabo en el grupo (Fernández de las Heras, van der Geize *et al.* 2012) sirven de piedra angular para el diseño mediante Ingeniería Metabólica de cepas capaces de producir compuestos tales como ADD o 9OH-AD a partir de AD en *R. ruber*. Además, la cepa de *R. ruber* que tiene bloqueada las actividades enzimáticas que procesan AD podría servir como punto de partida para su utilización como una biofactoría; la incorporación en *trans* de un actividad enzimática que introdujera modificaciones en la molécula de AD permitiría obtener otros intermediarios de interés (Figura 2A). Por ejemplo, la incorporación en este mutante de *R. ruber* de la 17β-Hidroxiesteroide deshidrogenasa (EC 1.1.1.51) permitiría la acumulación de testosterona a partir de AD. Esta estrategia ha sido llevada a cabo con éxito en *M. smegmatis* en la que se ha obtenido testosterona a partir de esteroles naturales (Fernández-Cabezón, Galán *et al.* 2016). En el caso de *R. ruber* sería necesario previamente la construcción de un vector de expresión que permita desarrollar todas estas posibilidades biotecnológicas.

Por otra parte, la obtención de intermediarios a partir de colesterol o fitoesteroles en cepas de *Rhodococcus* necesitará de más estudios para poder implementarse.

VIII. CONCLUSIONES

CONCLUSIONES

El trabajo descrito en esta Tesis Doctoral ha dado lugar a las siguientes conclusiones:

1. El genoma de *Rhodococcus ruber* Chol-4 muestra multiplicidad génica lo que le dota de versatilidad metabólica y le permite usar una amplia variedad de compuestos aromáticos y esteroides como fuente de carbono y energía.

2. Se ha confirmado la existencia de siete isoenzimas involucradas en la apertura del anillo esteroideo A/B, tres 3-cetoesteroides-Δ1-deshidrogenasas (KstD), tres 3-cetoesteroide 9α-hidroxilasas subunidad KshA y una unidad KshB.

3. La caracterización funcional in vitro de las tres KstDs muestra que KstD2 y probablemente KstD1 también están relacionadas con el catabolismo del AD mientras que KstD2 y KstD3 parecen estar relacionadas con el metabolismo del colesterol. KstD2 tiene actividad en un amplio rango de sustratos, por el contrario, KstD3 tiene un rango de sustratos más limitado principalmente a esteroides saturados.

4. En cuanto a funcionalidad biológica de las KshAs, se ha demostrado que la enzima KshA1 puede sustituir funcionalmente la actividad de KshA2 y KshA3 en presencia de AD y colesterol, respectivamente. Sin embargo es indispensable para la degradación del colato. También, que KshA2 está involucrada en la degradación de esteroles de cadena corta, mientras que la KshA3 está relacionada con la degradación de esteroles de cadena larga. La expresión de los genes kstDs y kshAs es basal, pero también presenta altos niveles de inducción adicional dependiendo de la fuente de carbono accesible, demostrando un mecanismo de regulación complejo.

5. Los resultados de esta Tesis Doctoral apoyan la hipótesis de que la degradación del colesterol ocurre mayoritariamente por una vía alternativa a la del AD y que los intermediarios de la degradación del colesterol no pasan necesariamente por el AD, el ADD o el 9OH-AD en Rhodococcus ruber Chol-4. También los resultados indican que la degradación de la cadena lateral de esteroides es independiente de la ruptura del anillo.

6. Se propone al género Rhodococcus como un candidato a biofactoría celular en la producción de compuestos deshidrogenados en C1-C2 a partir de diferentes esteroides.

IX. BIBLIOGRAFÍA

BIBLIOGRAFÍA

Abul-Hajj, Y. J. (1978). "Isolation of vitamin K2 (35) from *Nocardia restrictus* and *Corynebacterium simplex*. A natural electron acceptor in microbial steroid ring A dehydrogenations." J Biol Chem 253(7): 2356-2360.

Aherne, G. and R. Briggs (1989). "The relevance of the presence of certain synthetic steroids in the aquatic environment." Journal of Pharmacy and Pharmacology 41(10): 735-736.

Ahmad, S., P. Roy, S. Basu and B. Johri (1993). "Cholesterol side-chain cleavage by immobilized cells of *Rhodococcus equi* DSM 89-133." Indian journal of experimental biology 31(4): 319-322.

Ahn, K. W. and N. S. Sampson (2004). "Cholesterol oxidase senses subtle changes in lipid bilayer structure." Biochemistry 43(3): 827-836.

Aihara, H., K. Watanabe and R. Nakamura (1988). "Degradation of cholesterol in egg yolk by *Rhodococcus equi* No. 23." Journal of Food science 53(2): 659-660.

Alvarez, H. M. (2010). Biology of *Rhodococcus*. Berlin Heidelberg, Springer-Verlag.

Álvarez, H. M. (2010). Central Metabolism of Species of the Genus *Rhodococcus.* Berlin Heidelberg, Springer-Verlag.

Amoroso, M. J., C. S. Benimeli and S. A. Cuozzo (2013). Actinobacteria: Application in Bioremediation and Production of Industrial Enzymes. Florida, USA, CRC Press.

Andor, A., A. Jekkel, D. A. Hopwood, F. Jeanplong, É. Ilkőy, A. Kónya, I. Kurucz and G. Ambrus (2006). "Generation of useful insertionally blocked sterol degradation pathway mutants of fast-growing mycobacteria and cloning, characterization, and expression of the terminal oxygenase of the 3-ketosteroid 9α-hydroxylase in *Mycobacterium smegmatis* mc2155." Applied and environmental microbiology 72(10): 6554-6559.

Anesiadis, N., W. R. Cluett and R. Mahadevan (2008). "Dynamic metabolic engineering for increasing bioprocess productivity." Metabolic engineering 10(5): 255-266.

Arias-Barrau, E., E. R. Olivera, J. M. Luengo, C. Fernandez, B. Galan, J. L. Garcia, E. Diaz and B. Minambres (2004). "The homogentisate pathway: a central catabolic pathway involved in the degradation of L-phenylalanine, L-tyrosine, and 3-hydroxyphenylacetate in *Pseudomonas putida*." J Bacteriol 186(15): 5062-5077.

Aries, V., P. Goddard and M. Hill (1971). "Degradation of steroids by intestinal bacteria. III. 3-oxo-5β-steroid Δ1-dehydrogenase and 3-oxo-5β-steroid Δ4-dehydrogenase." Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism 248(3): 482-488.

Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke and O. Zagnitko (2008). "The RAST Server: rapid annotations using subsystems technology." BMC Genomics 9: 75.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M.

A. Alekseyev and P. A. Pevzner (2012). "SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing." J Comput Biol 19(5): 455-477.

Barel-Cohen, K., L. S. Shore, M. Shemesh, A. Wenzel, J. Mueller and N. Kronfeld-Schor (2006). "Monitoring of natural and synthetic hormones in a polluted river." J Environ Manage 78(1): 16-23.

Barry, S. M. and G. L. Challis (2013). "Mechanism and catalytic diversity of rieske non-heme iron-dependent oxygenases." ACS catalysis 3(10): 2362-2370.

Bashyam, M. D., D. Kaushal, S. K. Dasgupta and A. K. Tyagi (1996). "A study of mycobacterial transcriptional apparatus: identification of novel features in promoter elements." J Bacteriol 178(16): 4847-4853.

Batie, C. J., E. LaHaie and D. P. Ballou (1987). "Purification and characterization of phthalate oxygenase and phthalate oxygenase reductase from *Pseudomonas cepacia*." J Biol Chem 262(4): 1510-1518.

Bell, K. S., J. C. Philp, D. W. Aw and N. Christofi (1998). "The genus *Rhodococcus*." J Appl Microbiol 85(2): 195-210.

Bergstrand, L. H., E. Cardenas, J. Holert, J. D. Van Hamme and W. W. Mohn (2016). "Delineation of Steroid-Degrading Microorganisms through Comparative Genomic Analysis." MBio 7(2).

Bhatti, H. N. and R. A. Khera (2012). "Biological transformations of steroidal compounds: a review." Steroids 77(12): 1267-1290.

Bigey, F., B. Grossiord, C. C. K. Chion, A. Arnaud and P. Galzy (1995). "Brevibacterium linens pBL33 and *Rhodococcus rhodochrous* pRC1 cryptic plasmids replicate in Rhodococcus sp. R312 (formerly Brevibacterium sp. R312)." Gene 154(1): 77-79.

Bloch, K. (1965). "The biological synthesis of cholesterol." Science 150: 19-28.

Bode, H. B., B. Zeggel, B. Silakowski, S. C. Wenzel, H. Reichenbach and R. Müller (2003). "Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2, 3 (S)-oxidosqualene cyclase from the myxobacterium *Stigmatella aurantiaca*." Molecular microbiology 47(2): 471-481.

Bode, N. J. and K. H. Darwin (2014). "The Pup-Proteasome System of Mycobacteria." Microbiol Spectr 2(5).

Botelho, A., A. Canto, C. Leao and M. V. Cunha (2015). "Clustered regularly interspaced short palindromic repeats (CRISPRs) analysis of members of the *Mycobacterium tuberculosis* complex." Methods Mol Biol 1247: 373-389.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Anal. Biochem. 72: 248-254.

Bragin, E. Y., V. Y. Shtratnikova, D. Dovbnya, M. Schelkunov, Y. A. Pekov, S. Malakho, O. Egorova, T. Ivashina, S. Sokolov and V. Ashapkin (2013). "Comparative analysis of genes encoding key steroid core oxidation enzymes in fast-growing *Mycobacterium spp*. strains." The Journal of steroid biochemistry and molecular biology 138: 41-53.

Brunhuber, N. M., J. B. Thoden, J. S. Blanchard and J. L. Vanhooke (2000). "*Rhodococcus* L-phenylalanine dehydrogenase: kinetics, mechanism, and structural basis for catalytic specificity." Biochemistry 39(31): 9174-9187.

Brzostek, A., B. Dziadek, A. Rumijowska-Galewicz, J. Pawelczyk and J. Dziadek (2007). "Cholesterol oxidase is required for virulence of *Mycobacterium tuberculosis*." FEMS Microbiol Lett 275(1): 106-112.

Brzostek, A., T. Sliwinski, A. Rumijowska-Galewicz, M. Korycka-Machala and J. Dziadek (2005). "Identification and targeted disruption of the gene encoding the main 3-ketosteroid dehydrogenase in *Mycobacterium smegmatis*." Microbiology 151(Pt 7): 2393-2402.

Buhaescu, I. and H. Izzedine (2007). "Mevalonate pathway: a review of clinical and therapeutical implications." Clinical biochemistry 40(9): 575-584.

Byrne, G. A., D. A. Russell, X. Chen and W. G. Meijer (2007). "Transcriptional regulation of the virR operon of the intracellular pathogen *Rhodococcus equi*." Journal of bacteriology 189(14): 5082-5089.

Capyk, J. K., I. Casabon, *et al.* (2011). "Activity of 3-ketosteroid 9alpha-hydroxylase (KshAB) indicates cholesterol side chain and ring degradation occur simultaneously in *Mycobacterium tuberculosis*." J Biol Chem 286(47): 40717-40724.

Capyk, J. K., I. D'Angelo, N. C. Strynadka and L. D. Eltis (2009). "Characterization of 3-ketosteroid 9{alpha}-hydroxylase, a rieske oxygenase in the cholesterol degradation pathway of *Mycobacterium tuberculosis*." J Biol Chem 284(15): 9937-9946.

Capyk, J. K., N. C. Strynadka and L. D. Eltis (2010). "3-Ketosteroid-9α-Hydroxylase." Encyclopedia of Inorganic and Bioinorganic Chemistry.

Capyk, J. K., R. Kalscheuer, G. R. Stewart, J. Liu, H. Kwon, R. Zhao, S. Okamoto, W. R. Jacobs, L. D. Eltis and W. W. Mohn (2009). "Mycobacterial cytochrome p450 125 (cyp125) catalyzes the terminal hydroxylation of c27 steroids." Journal of Biological Chemistry 284(51): 35534-35542.

Casabon, I., A. M. Crowe, J. Liu and L. D. Eltis (2013). "FadD3 is an acyl-CoA synthetase that initiates catabolism of cholesterol rings C and D in actinobacteria." Molecular microbiology 87(2): 269-283.

Casabon, I., S. H. Zhu, H. Otani, J. Liu, W. W. Mohn and L. D. Eltis (2013). "Regulation of the KstR2 regulon of *Mycobacterium tuberculosis* by a cholesterol catabolite." Molecular microbiology 89(6): 1201-1212.

Chang, H., Y. Wan, S. Wu, Z. Fan and J. Hu (2011). "Occurrence of androgens and progestogens in wastewater treatment plants and receiving river waters: comparison to estrogens." Water Res 45(2): 732-740.

Chen, M. M., F. Q. Wang, L. C. Lin, K. Yao and D. Z. Wei (2012). "Characterization and application of fusidane antibiotic biosynthesis enzyme 3-ketosteroid-(1)-dehydrogenase in steroid transformation." Appl Microbiol Biotechnol. In press.

Chiang, Y. R., J. Y. Fang, *et al.* (2010). "Initial steps in anoxic testosterone degradation by Steroidobacter denitrificans." Microbiology 156(Pt 7): 2253-2259.

Chiba, K., Y. Hoshino, K. Ishino, T. Kogure, Y. Mikami, Y. Uehara and J. Ishikawa (2007). "Construction of a pair of practical *Nocardia-Escherichia* coli shuttle vectors." Jpn. J. Infect. Dis. 60(1): 45-47.

Chipley, J. R., M. Dreyfuss and R. Smucker (1975). "Cholesterol metabolism by *Mycobacterium*." Microbios 12: 199-207.

Choi, K.-P., I. Molnar, M. Yamashita and Y. Murooka (1995). "Purification and Characterization of the 3-Ketosteroid-Δ1-Dehydrogenase of *Arthrobacter simplex* Produced in *Streptomyces liuidans*." The Journal of Biochemistry 117(5): 1043-1049.

Coulombe, R., K. Q. Yue, S. Ghisla and A. Vrielink (2001). "Oxygen access to the active site of cholesterol oxidase through a narrow channel is gated by an Arg-Glu pair." J Biol Chem 276(32): 30435-30441.

Crowe, A. M., P. J. Stogios, I. Casabon, E. Evdokimova, A. Savchenko and L. D. Eltis (2015). "Structural and functional characterization of a ketosteroid transcriptional regulator of *Mycobacterium tuberculosis*." J Biol Chem 290(2): 872-882.

Dabrock, B., M. Kesseler, B. Averhoff and G. Gottschalk (1994). "Identification and characterization of a transmissible linear plasmid from *Rhodococcus erythropolis* BD2 that encodes isopropylbenzene and trichloroethene catabolism." Appl Environ Microbiol 60(3): 853-860.

Dahl, J. S., C. E. Dahl and K. Bloch (1982). "Role of membrane sterols in *Mycoplasma capricolum*." Review of Infectious Diseases 4(Supplement 1): S93-S96.

De Carvalho, C. C. and M. M. R. Da Fonseca (2005). "The remarkable *Rhodococcus erythropolis*." Applied microbiology and biotechnology 67(6): 715-726.

de Jong, A., H. Pietersma, M. Cordes, O. P. Kuipers and J. Kok (2012). "PePPER: a webserver for prediction of prokaryote promoter elements and regulons." BMC Genomics 13: 299.

De Mot, R., I. Nagy, A. De Schrijver, P. Pattanapipitpaisal, G. Schoofs and J. Vanderleyden (1997). "Structural analysis of the 6 kb cryptic plasmid pFAJ2600 from *Rhodococcus erythropolis* NI86/21 and construction of *Escherichia coli-Rhodococcus* shuttle vectors." Microbiology 143(10): 3137-3147.

de Vries, J. and W. Wackernagel (2002). "Integration of foreign DNA during natural transformation of *Acinetobacter sp.* by homology-facilitated illegitimate recombination." Proc Natl Acad Sci U S A 99(4): 2094-2099.

Denis-Larose, C., D. Labbe, H. Bergeron, A. M. Jones, C. W. Greer, J. Al-Hawari, M. J. Grossman, B. M. Sankey and P. Lau (1997). "Conservation of plasmid-encoded dibenzothiophene desulfurization genes in several rhodococci." Applied and environmental microbiology 63(7): 2915-2919.

Denome, S. A. and K. D. Young (1995). "Identification and activity of two insertion sequence elements in *Rhodococcus sp.* strain IGTS8." Gene 161(1): 33-38.

Di Gennaro, P., P. Terreni, G. Masi, S. Botti, F. De Ferra and G. Bestetti (2010). "Identification and characterization of genes involved in naphthalene degradation in *Rhodococcus opacus* R7." Appl Microbiol Biotechnol 87(1): 297-308.

Dijkstra, B. W., N. van Oosterwijk and A. Rohman (2016). "Structure and Catalytic Mechanism of 3-Ketosteroid Dehydrogenases." Procedia Chemistry 18: 3-11.

Ding, J. Y., J. H. Shiu, *et al.* (2016). "Genomic Insight into the Host-Endosymbiont Relationship of *Endozoicomonas montiporae* CL-33(T) with its Coral Host." Front Microbiol 7: 251. doi: 10.3389/fmicb.2016.00251.

Döhr, O., M. J. Paine, T. Friedberg, G. C. Roberts and C. R. Wolf (2001). "Engineering of a functional human NADH-dependent cytochrome P450 system." Proceedings of the National Academy of Sciences 98(1): 81-86.

Donova, M. V. and O. V. Egorova (2012). "Microbial steroid transformations: current state and prospects." Appl Microbiol Biotechnol 94(6): 1423-1447.

Doukyu, N. (2009). "Characteristics and biotechnological applications of microbial cholesterol oxidases." Appl Microbiol Biotechnol 83(5): 825-837.

Dresen, C., L. Y. Lin, *et al.* (2010). "A flavin-dependent monooxygenase from *Mycobacterium tuberculosis* involved in cholesterol catabolism." J. Biol. Chem. 285(29): 22264-22275.

Driscoll, M. D., K. J. McLean, C. Levy, N. Mast, I. A. Pikuleva, P. Lafite, S. E. Rigby, D. Leys and A. W. Munro (2010). "Structural and biochemical characterization of *Mycobacterium tuberculosis* CYP142 evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen." Journal of Biological Chemistry 285(49): 38270-38282.

Drzyzga, O., L. Fernandez de las Heras, V. Morales, J. M. Navarro Llorens and J. Perera (2011). "Cholesterol degradation by *Gordonia cholesterolivorans*." Appl Environ Microbiol 77(14): 4802-4810.

Duran, R. (1998). "New shuttle vectors for *Rhodococcus sp.* R312 (formerly Brevibacterium sp. R312), a nitrile hydratase producing strain." Journal of basic microbiology 38(2): 101-106.

Durante-Rodriguez, G., V. de Lorenzo and E. Martinez-Garcia (2014). "The Standard European Vector Architecture (SEVA) plasmid toolkit." Methods Mol Biol 1149: 469-478.

Dym, O. and D. Eisenberg (2001). "Sequence-structure analysis of FAD-containing proteins." Protein Science 10(9): 1712-1728.

Eaton, R. W., O. V. Selifonova and R. M. Gedney (1998). "Isopropylbenzene catabolic pathway in *Pseudomonas putida* RE204: nucleotide sequence analysis of the ipb operon and neighboring DNA from pRE4." Biodegradation 9(2): 119-132.

Embley, T. and E. Stackebrandt (1994). "The molecular phylogency and systematics of the actinomycetes." Annual Reviews in Microbiology 48(1): 257-289.

Emenike, C., P. Agamuthu and S. Fauziah (2016). "*Blending Bacillus sp., Lysinibacillus sp*. and *Rhodococcus sp.* for optimal reduction of heavy metals in leachate contaminated soil." Environmental Earth Sciences 75(1): 26.

Fahy, E., S. Subramaniam, H. A. Brown, C. K. Glass, A. H. Merrill, R. C. Murphy, C. R. Raetz, D. W. Russell, Y. Seyama and W. Shaw (2005). "A comprehensive classification system for lipids." Journal of lipid research 46(5): 839-862.

Fernandes, P., A. Cruz, *et al.* (2003). "Microbial conversion of steroid compounds: recent developments." Enzyme Microbial Technology 32: 688-705.

Fernández-Cabezón, L., B. Galán and J. L. García (2016). "Engineering *Mycobacterium smegmatis* for testosterone production." Microbial Biotechnology.

Fernández-Cabezón, L., E. García-Fernández, B. Galán and J. L. García (2017). "Molecular characterization of a new gene cluster for steroid degradation in *Mycobacterium smegmatis*." Environmental Microbiology.

Fernández de las Heras, L. (2013). Estudio genético y bioquímico de la ruta de degradación del colesterol en *Rhodococcus spp.*, Universidad Complutense de Madrid.

Fernández de las Heras, L., E. García Fernández, J. María Navarro Llorens, J. Perera and O. Drzyzga (2009). "Morphological, physiological, and molecular characterization of a newly isolated steroid-degrading actinomycete, identified as *Rhodococcus ruber* strain Chol-4." Curr. Microbiol. 59(5): 548-553.

Fernández de las Heras, L., J. Perera and J. M. N. Llorens (2014). "Cholesterol to cholestenone oxidation by ChoG, the main extracellular cholesterol oxidase of *Rhodococcus ruber* strain Chol-4." The Journal of steroid biochemistry and molecular biology 139: 33-44.

Fernández de las Heras, L., R. van der Geize, O. Drzyzga, J. Perera and J. M. N. Llorens (2012). "Molecular characterization of three 3-ketosteroid-Δ 1-dehydrogenase isoenzymes of *Rhodococcus ruber* strain Chol-4." The Journal of steroid biochemistry and molecular biology 132(3): 271-281.

Fernández de las Heras, L., S. Alonso, A. de la Vega de Leon, D. Xavier, J. Perera and J. M. Navarro Llorens (2013). "Draft Genome Sequence of the Steroid Degrader *Rhodococcus ruber* Strain Chol-4." Genome Announc. 1(3).

Fernández de las Heras, L., V. Mascaraque, E. García Fernandez, J. M. Navarro-Llorens, J. Perera and O. Drzyzga (2011). "ChoG is the main inducible extracellular cholesterol oxidase of *Rhodococcus sp.* strain CECT3014." Microbiol. Res. 166(5): 403-418.

Ferraro, D. J., L. Gakhar and S. Ramaswamy (2005). "Rieske business: structure-function of Rieske non-heme oxygenases." Biochem Biophys Res Commun 338(1): 175-190.

Finnerty, W. R. (1992). "The biology and genetics of the genus *Rhodococcus.*" Annu Rev Microbiol 46: 193-218.

Florin, C., T. Kohler, M. Grandguillot and P. Plesiat (1996). "*Comamonas testosteroni* 3-ketosteroid-delta 4(5 alpha)-dehydrogenase: gene and protein characterization." J Bacteriol 178(11): 3322-3330.

Francis, I., A. De Keyser, P. De Backer, C. Simon-Mateo, J. Kalkus, I. Pertry, W. Ardiles-Diaz, R. De Rycke, O. M. Vandeputte, M. El Jaziri, M. Holsters and D. Vereecke (2012). "pFiD188, the linear virulence plasmid of *Rhodococcus fascians* D188." Mol Plant Microbe Interact 25(5): 637-647.

Galán, B., I. Uhía, E. García-Fernández, I. Martínez, E. Bahíllo, J. L. Barredo, L. Fernández-Cabezón and J. L. García (2016). "*Mycobacterium smegmatis* is a suitable cell factory for the production of steroidic synthons." Microbial Biotechnology.

García Fernández, J. (2015). "Estudio del transporte de esteroides en *Mycobacterium smegmatis* mc2155 y sus aplicaciones biotecnológicas."

García, J. L., I. Uhía and B. Galan (2012). "Catabolism and biotechnological applications of cholesterol degrading bacteria." Microb. Biotechnol. 5(6): 679-699.

García-Fernández, E., D. J. Frank, B. Galán, P. M. Kells, L. M. Podust, J. L. García and P. R. Ortiz de Montellano (2013). "A highly conserved mycobacterial cholesterol catabolic pathway." Environmental microbiology 15(8): 2342-2359.

García-Fernández, E., F. J. Medrano, B. Galán and J. L. García (2014). "Deciphering The Transcriptional Regulation of Cholesterol Catabolic Pathway in *Mycobacteria*: Identification of the Inducer of KstR Repressor." J Biol Chem 289: 17576-17588.

García-Fernández, J., B. Galán, F. J. Medrano and J. L. García (2015). "Characterization of the KstR2 regulator responsible of the lower cholesterol degradative pathway in *Mycobacterium smegmatis*." Environmental microbiology reports 7(1): 155-163.

Gebhardt, H., X. Meniche, M. Tropis, R. Krämer, M. Daffe and S. Morbach (2007). "The key role of the mycolic acid content in the functionality of the cell wall permeability barrier in *Corynebacterineae*." Microbiology 153(5): 1424-1434.

Goncalves, E. R., H. Hara, D. Miyazawa, J. E. Davies, L. D. Eltis and W. W. Mohn (2006). "Transcriptomic assessment of isozymes in the biphenyl pathway of *Rhodococcus sp.* strain RHA1." Appl Environ Microbiol 72(9): 6183-6193.

Goodfellow, M. (1989). Genus *Rhodococcus.* Bergey's Manual of Systematic Bacteriology. S. T. Williams, M. E. Sharpe and J. G. and Holt. Baltimore, Williams and Wilkins. vol. 4: pp. 2362-2371.

Goodfellow, M., G. Alderson and J. Chun (1998). "Rhodococcal systematics: problems and developments." Antonie van Leeuwenhoek 74(1-3): 3-20.

Goswami, L., M. T. Namboodiri, R. V. Kumar, K. Pakshirajan and G. Pugazhenthi (2017). "Biodiesel production potential of oleaginous *Rhodococcus opacus* grown on biomass gasification wastewater." Renewable Energy 105: 400-406.

Gracia, T., P. D. Jones, E. B. Higley, K. Hilscherova, J. L. Newsted, M. B. Murphy, A. K. Chan, X. Zhang, M. Hecker, P. K. Lam, R. S. Wu and J. P. Giesy (2008). "Modulation of steroidogenesis by coastal waters and sewage effluents of Hong Kong, China, using the H295R assay." Environ Sci Pollut Res Int 15(4): 332-343.

Griffin, J. E., J. D. Gawronski, M. A. Dejesus, T. R. Ioerger, B. J. Akerley and C. M. Sassetti (2011). "High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism." PLoS Pathog 7(9): e1002251.

Groh, H., R. Komel, V. Deppmeyer, W. Schade and C. Hörhold (1980). "Steroid transforming enzymes from microorganisms: the reverse reaction of the steroid-1-dehydrogenase from *Nocardia*." Journal of steroid biochemistry 13(12): 1413-1415.

Grummt, I., A. Kuhn, I. Bartsch and H. Rosenbauer (1986). "A transcription terminator located upstream of the mouse rDNA initiation site affects rRNA synthesis." Cell 47(6): 901-911.

Guillette, L., A. A. Rooney, D. A. Crain and E. F. Orlando (1999). Steroid hormones as biomarkers of endocrine disruption in wildlife. Environmental Toxicology and Risk Assessment: Standardization of Biomarkers for Endocrine Disruption and Environmental Assessment: 8th Volume, ASTM International.

Gust, B., G. L. Challis, K. Fowler, T. Kieser and K. F. Chater (2003). "PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin." Proc Natl Acad Sci U S A 100(4): 1541-1546.

Hall, T. A. (1999). "BioEdit a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT." Nucl. Acids Symp. Ser. 41: 95-98.

Hashimoto, Y., M. Nishiyama, F. Yu, I. Watanabe, S. Horinouchi and T. Beppu (1992). "Development of a host-vector system in a *Rhodococcus* strain and its use for expression of the cloned nitrile hydratase gene cluster." J Gen Microbiol 138(5): 1003-1010.

He, Y., X. Li, H. Ben, X. Xue and B. Yang (2017). "Lipids Production from Dilute Alkali Corn Stover Lignin by *Rhodococcus* Strains." ACS Sustainable Chemistry & Engineering.

Henderson, S. and B. Sollner-Webb (1986). "A transcriptional terminator is a novel element of the promoter of the mouse ribosomal RNA gene." Cell 47(6): 891-900.

Henderson, S. L., K. Ryan and B. Sollner-Webb (1989). "The promoter-proximal rDNA terminator augments initiation by preventing disruption of the stable transcription complex caused by polymerase read-in." Genes Dev 3(2): 212-223.

Hernandez, M. A., W. W. Mohn, E. Martínez, E. Rost, A. F. Alvarez and H. M. Alvarez (2008). "Biosynthesis of storage compounds by *Rhodococcus jostii* RHA1 and global identification of genes involved in their metabolism." BMC Genomics 9: 600.

Hilyard, E. J., J. M. Jones-Meehan, B. J. Spargo and R. T. Hill (2008). "Enrichment, isolation, and phylogenetic identification of polycyclic aromatic hydrocarbon-degrading bacteria from Elizabeth River sediments." Appl Environ Microbiol 74(4): 1176-1182.

Hirasawa, K., Y. Ishii, M. Kobayashi, K. Koizumi and K. Maruhashi (2001). "Improvememt of Desulfurization Activity in *Rhodococcus erythropolis* KA2-5-1 by Genetic Engineering." Bioscience, biotechnology, and biochemistry 65(2): 239-246.

Ho, N. A., S. S. Dawes, A. M. Crowe, I. Casabon, C. Gao, S. L. Kendall, E. N. Baker, L. D. Eltis and J. S. Lott (2016). "The Structure of the Transcriptional Repressor KstR in Complex with CoA Thioester Cholesterol Metabolites Sheds Light on the Regulation of Cholesterol Catabolism in *Mycobacterium tuberculosis*." J Biol Chem 291(14): 7256-7266.

Hofmann, A. and L. Hagey (2008). "Bile acids: chemistry, pathochemistry, biology, pathobiology, and therapeutics." Cellular and Molecular Life Sciences 65(16): 2461-2483.

Hondalus, M. K. (1997). "Pathogenesis and virulence of *Rhodococcus equi*." Vet Microbiol 56(3-4): 257-268.

Hopwood, D. A., M. J. Bibb, K. F. Chater, T. Kieser, C. J. Bruton, H. M. Kieser, D. J. Lydiate, C. P. Smith, J. M. Ward and H. Schrempf, Eds. (1985). Genetic Manipulation of *Streptomyces.* A Laboratory Manual., Norwich : John Innes Foundation.

Hori, K., A. Kobayashi, H. Ikeda and H. Unno (2009). "*Rhodococcus aetherivorans* IAR1, a new bacterial strain synthesizing poly(3-hydroxybutyrate-co-3-hydroxyvalerate) from toluene." J Biosci Bioeng 107(2): 145-150.

Horinouchi, M., T. Hayashi and T. Kudo (2012). "Steroid degradation in *Comamonas testosteroni*." J Steroid Biochem Mol Biol 129(1-2): 4-14.

Horinouchi, M., T. Hayashi, *et al.* (2003). "A new bacterial steroid degradation gene cluster in *Comamonas testosteroni* TA441 which consists of aromatic-compound degradation genes for seco-steroids and 3-ketosteroid dehydrogenase genes." Appl. Environ. Microbiol. 69(8): 4421-4430.

Horinouchi, M., T. Hayashi, H. Koshino, T. Kurita and T. Kudo (2005). "Identification of 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid, 4-hydroxy-2-oxohexanoic acid, and 2-hydroxyhexa-2,4-dienoic acid and related enzymes involved in testosterone degradation in Comamonas testosteroni TA441." Appl Environ Microbiol 71(9): 5275-5281.

Horinouchi, M., T. Hayashi, T. Yamamoto and T. Kudo (2003). "A new bacterial steroid degradation gene cluster in *Comamonas testosteroni* TA441 which consists of aromatic-compound degradation genes for seco-steroids and 3-ketosteroid dehydrogenase genes." Appl Environ Microbiol 69(8): 4421-4430.

Horinouchi, M., T. Kurita, *et al.* (2010). "Steroid degradation genes in *Comamonas testosteroni* TA441: Isolation of genes encoding a Delta4(5)-isomerase and 3alpha- and 3beta-dehydrogenases and evidence for a 100 kb steroid degradation gene hot spot." J. Steroid Biochem. Mol. Biol. 122(4): 253-263.

Hosek, J., P. Svastova, M. Moravkova, I. Pavlik and M. Bartos (2006). "Methods of mycobacterial DNA isolation from different biological material: a review." Veterinary Research 524: 180-192.

Hu, Y., R. van der Geize, *et al.* (2010). "3-Ketosteroid 9alpha-hydroxylase is an essential factor in the pathogenesis of *Mycobacterium tuberculosis*." Mol Microbiol 75(1): 107-121.

Hu, Y., R. Van Der Geize, G. S. Besra, S. S. Gurcha, A. Liu, M. Rohde, M. Singh and A. Coates (2010). "3-ketosteroid 9α-hydroxylase is an essential factor in the pathogenesis of *Mycobacterium tuberculosis*." Molecular microbiology 75(1): 107-121.

Hughes, J., Y. C. Armitage and K. C. Symes (1998). "Application of whole cell rhodococcal biocatalysts in acrylic polymer manufacture." Antonie van Leeuwenhoek 74(1-3): 107-118.

Iida, T., T. Waki, K. Nakamura, Y. Mukouzaka and T. Kudo (2009). "The GAF-like-domain-containing transcriptional regulator DfdR is a sensor protein for dibenzofuran and several hydrophobic aromatic compounds." Journal of bacteriology 191(1): 123-134.

Itagaki, E., T. Hatta, T. Wakabayashi and K. Suzuki (1990). "Spectral properties of 3-ketosteroid-delta 1-dehydrogenase from *Nocardia corallina*." Biochim Biophys Acta 1040(2): 281-286.

Itagaki, E., T. Wakabayashi and T. Hatta (1990). "Purification and characterization of 3-ketosteroid-delta 1-dehydrogenase from Nocardia corallina." Biochim Biophys Acta 1038(1): 60-67.

Ivashina, T. V., V. M. Nikolayeva, D. V. Dovbnya and M. V. Donova (2012). "Cholesterol oxidase ChoD is not a critical enzyme accounting for oxidation of sterols to 3-keto-4-ene steroids in fast-growing *Mycobacterium sp.* VKM Ac-1815D." J Steroid Biochem Mol Biol 129(1-2): 47-53.

Ivshina, I., V. Grishko, E. Nogovitsina, T. Kukina and G. Tolstikov (2004). "Bioconversion of beta-sitosterol and its complex esters by *Rhodococcus actinobacteria*." Prikladnaia biokhimiia i mikrobiologiia 41(6): 626-633.

Iwabuchi, N., M. Sunairi, M. Urai, C. Itoh, H. Anzai, M. Nakajima and S. Harayama (2002). "Extracellular polysaccharides of *Rhodococcus rhodochrous* S-2 stimulate the degradation of aromatic components in crude oil by indigenous marine bacteria." Appl Environ Microbiol 68(5): 2337-2343.

JCBN (1989). "The nomenclature of steroids. Recommendations 1989." European journal of biochemistry 186(3): 429-458.

Jobst, B., K. Schuhle, U. Linne and J. Heider (2010). "ATP-dependent carboxylation of acetophenone by a novel type of carboxylase." J Bacteriol 192(5): 1387-1394.

Jones, O., N. Voulvoulis and J. Lester (2001). "Human pharmaceuticals in the aquatic environment a review." Environmental technology 22(12): 1383-1394.

Kalkus, J., R. Menne, M. Reh and H. G. Schlegel (1998). "The terminal structures of linear plasmids from *Rhodococcus opacus.*"

Kannenberg, E. L. and K. Poralla (1999). "Hopanoid biosynthesis and function in bacteria." Naturwissenschaften 86(4): 168-176.

Kendall, S. L., M. Withers, C. N. Soffair, N. J. Moreland, S. Gurcha, B. Sidders, R. Frita, A. Ten Bokum, G. S. Besra, J. S. Lott and N. G. Stoker (2007). "A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*." Mol Microbiol 65(3): 684-699.

Kendall, S. L., P. Burgess, R. Balhana, M. Withers, A. Ten Bokum, J. S. Lott, C. Gao, I. Uhía-Castro and N. G. Stoker (2010). "Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2." Microbiology 156(Pt 5): 1362-1371.

Kesseler, M., E. R. Dabbs, B. Averhoff and G. Gottschalk (1996). "Studies on the isopropylbenzene 2,3-dioxygenase and the 3-isopropylcatechol 2,3-dioxygenase genes encoded by the linear plasmid of *Rhodococcus erythropolis* BD2." Microbiology 142 (Pt 11): 3241-3251.

Kieslich, K. (1985). "Microbial side-chain degradation of sterols." J Basic Microbiol 25: 461-474.

Kisiela, M., A. Skarka, B. Ebert and E. Maser (2012). "Hydroxysteroid dehydrogenases (HSDs) in bacteria - A bioinformatic perspective." J Steroid Biochem Mol Biol 129(1-2): 31-46.

Kitagawa, W. and T. Tamura (2008). "A quinoline antibiotic from *Rhodococcus erythropolis* JCM 6824." Journal of antibiotics 61(11): 680.

Klein, U., G. Gimpl and F. Fahrenholz (1995). "Alteration of the myometrial plasma membrane cholesterol content with b-cyclodextrin modulates the binding affinity of the oxytocin receptor." Biochemistry 34(42): 13784-13793.

Klepp, L. I., M. A. Forrellad, A. V. Osella, F. C. Blanco, E. J. Stella, M. V. Bianco, M. D. Santangelo, C. Sassetti, M. Jackson, A. A. Cataldi, F. Bigi and H. R. Morbidoni (2012). "Impact of the deletion of the six mce operons in *Mycobacterium smegmatis*." Microbes and Infection 14(7-8): 590-599.

Knol, J. (2011). Natural diversity and catalytic mechanisms of Delta1- and Delta4-3ketosteroid dehydrogenases from actinomycetal origin. PhD thesis, Rijksuniversiteit Groningen.

Knol, J., K. Bodewits, G. I. Hessels, L. Dijkhuizen and R. van der Geize (2008). "3-Keto-5alpha-steroid Delta(1)-dehydrogenase from *Rhodococcus erythropolis* SQ1 and its orthologue in *Mycobacterium tuberculosis* H37Rv are highly specific enzymes that function in cholesterol catabolism." Biochem J 410(2): 339-346.

Koerner, R. J., M. Goodfellow and A. L. Jones (2009). "The genus Dietzia: a new home for some known and emerging opportunist pathogens." FEMS Immunology & Medical Microbiology 55(3): 296-305.

König, C., D. Eulberg, J. Gröning, S. Lakner, V. Seibert, S. R. Kaschabek and M. Schlömann (2004). "A linear megaplasmid, p1CP, carrying the genes for chlorocatechol catabolism of *Rhodococcus opacus* 1CP." Microbiology 150(9): 3075-3087.

Komeda, H., Y. Hori, M. Kobayashi and S. Shimizu (1996). "Transcriptional regulation of the *Rhodococcus rhodochrous* J1 nitA gene encoding a nitrilase." Proceedings of the National Academy of Sciences 93(20): 10572-10577.

Kostichka, K., L. Tao, M. Bramucci, J.-F. Tomb, V. Nagarajan and Q. Cheng (2003). "A small cryptic plasmid from *Rhodococcus erythropolis*: characterization and utility for gene expression." Applied microbiology and biotechnology 62(1): 61-68.

Kreit, J. and N. S. Sampson (2009). "Cholesterol oxidase: physiological functions." Febs Journal 276(23): 6844-6856.

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones and M. A. Marra (2009). "Circos: an information aesthetic for comparative genomics." Genome Res 19(9): 1639-1645.

Kulakov, L. A., G. J. Poelarends, D. B. Janssen and M. J. Larkin (1999). "Characterization of IS2112, a new insertion sequence from *Rhodococcus*, and its relationship with mobile elements belonging to the IS110 family." Microbiology 145 (Pt 3): 561-568.

Kulakova, A. N., K. A. Reid, M. J. Larkin, C. C. Allen and L. A. Kulakov (1996). "Isolation of *Rhodococcus rhodochrous* NCIMB13064 derivatives with new biodegradative abilities." FEMS microbiology letters 145(2): 227-231.

Kulakova, A. N., T. M. Stafford, M. J. Larkin and L. A. Kulakov (1995). "Plasmid pRTL1 controlling 1-chloroalkane degradation by *Rhodococcus rhodochrous* NCIMB13064." Plasmid 33(3): 208-217.

Kumari, L. and K. S. Shamsher (2015). "Cholesterol oxidase: Role in biotransformation of cholesterol." Journal of Applied Biology and Biotechnology Vol 3(06): 053-065.

Kuyukina, M. S. and I. B. Ivshina (2010). Application of *Rhodococcus* in bioremediation of contaminated environments. Biology of *Rhodococcus*, Springer: 231-262.

Kuyukina, M. S. and I. B. Ivshina (2010). *Rhodococcus* Biosurfactants: Biosynthesis, Properties, and Potential Applications. Berlin Heidelberg, Springer-Verlag.

Lamb, D. C., D. E. Kelly, N. J. Manning and S. L. Kelly (1998). "A sterol biosynthetic pathway in *Mycobacterium.*" FEBS Lett 437(1-2): 142-144.

Lario, P. I., N. Sampson and A. Vrielink (2003). "Sub-atomic resolution crystal structure of cholesterol oxidase: what atomic resolution crystallography reveals about enzyme mechanism and the role of the FAD cofactor in redox activity." J Mol Biol 326(5): 1635-1650.

Larkin, M. J., L. A. Kulakov and C. C. Allen (2005). "Biodegradation and *Rhodococcus*–masters of catabolic versatility." Current Opinion in Biotechnology 16(3): 282-290.

Larkin, M. J., L. A. Kulakov and C. C. R. Allen (2010). Genomes and Plasmids in *Rhodococcus*. Berlin Heidelberg, Springer-Verlag.

Larkin, M. J., R. De Mot, L. A. Kulakov and I. Nagy (1998). "Applied aspects of *Rhodococcus* genetics." Antonie van Leeuwenhoek 74(1-3): 133-153.

Le, R. K., T. Wells Jr, P. Das, X. Meng, R. J. Stoklosa, A. Bhalla, D. B. Hodge, J. S. Yuan and A. J. Ragauskas (2017). "Conversion of corn stover alkaline pre-treatment waste streams into biodiesel via Rhodococci." RSC Advances 7(7): 4108-4115.

Lee, S. and C. J. Sih (1967). "Mechanisms of Steroid Oxidation by Microorganisms. XII. Metabolism of Hexahydroindanpropionic Acid Derivatives*." Biochemistry 6(5): 1395-1403.

Lefebvre, P., B. Cariou, F. Lien, F. Kuipers and B. Staels (2009). "Role of bile acids and bile acid receptors in metabolic regulation." Physiological reviews 89(1): 147-191.

Lehmann, B. and M. Meurer (2010). "Vitamin D metabolism." Dermatologic therapy 23(1): 2-12.

Leng, D.-H., D.-X. Wang, J. Pan, Z.-T. Huang and M.-X. Wang (2009). "Highly efficient and enantioselective biotransformations of racemic azetidine-2-carbonitriles and their synthetic applications." The Journal of organic chemistry 74(16): 6077-6082.

Lessard, P. A., X. M. O'Brien, D. H. Currie and A. J. Sinskey (2004). "pB264, a small, mobilizable, temperature sensitive plasmid from *Rhodococcus*." BMC microbiology 4(1): 1.

Levy, H. R. and P. Talalay (1959). "Bacterial oxidation of steroids. II. Studies on the enzymatic mechanism of ring A dehydrogenation." J Biol Chem 234(8): 2014-2021.

Li, H. and R. Durbin (2010). "Fast and accurate long-read alignment with Burrows-Wheeler transform." Bioinformatics 26(5): 589-595.

Li, J., A. Vrielink, P. Brick and D. M. Blow (1993). "Crystal structure of cholesterol oxidase complexed with a steroid substrate: implications for flavin adenine dinucleotide dependent alcohol oxidases." Biochemistry 32: 11507-11515.

Li, Q., F. Ge, Y. Tan, G. Zhang and W. Li (2016). "Genome-Wide Transcriptome Profiling of *Mycobacterium smegmatis* MC2 155 Cultivated in Minimal Media Supplemented with Cholesterol, Androstenedione or Glycerol." International journal of molecular sciences 17(5): 689.

Liu, T. T., Y. Xu, H. Liu, S. Luo, Y. J. Yin, S. J. Liu and N. Y. Zhou (2011). "Functional characterization of a gene cluster involved in gentisate catabolism in *Rhodococcus sp.* strain NCIMB 12038." Appl Microbiol Biotechnol 90(2): 671-678.

Liu, Y., Y. Shen, Y. Qiao, L. Su, C. Li and M. Wang (2016). "The effect of 3-ketosteroid-Delta(1)-dehydrogenase isoenzymes on the transformation of AD to 9alpha-OH-AD by *Rhodococcus rhodochrous* DSM43269." J Ind Microbiol Biotechnol 43(9): 1303-1311.

Luengo, J. M., J. L. Garcia and E. R. Olivera (2001). "The phenylacetyl-CoA catabolon: a complex catabolic unit with broad biotechnological applications." Mol Microbiol 39(6): 1434-1442.

Lutton, C. (1991). "Dietary cholesterol, membrane cholesterol and cholesterol synthesis." Biochimie 73(10): 1327-1334.

Ma, Y. and H. Yu (2012). "Engineering of *Rhodococcus* cell catalysts for tolerance improvement by sigma factor mutation and active plasmid partition." J Ind Microbiol Biotechnol 39(10): 1421-1430.

Madigan, M. T., J. M. Martinko and J. Parker (1997). Brock biology of microorganisms, prentice hall Upper Saddle River, NJ.

Majewska, M. D. (2007). "Steroids and ion channels in evolution: from bacteria to synapses and mind. Evolutionary role of steroid regulation of GABA(A) receptors." Acta neurobiologiae experimentalis 67(3): 219-233.

Malaviya, A. and J. Gomes (2008). "Androstenedione production by biotransformation of phytosterols." Bioresour Technol 99(15): 6725-6737.

Martínková, L., B. Uhnakova, M. Patek, J. Nesvera and V. Kren (2009). "Biodegradation potential of the genus *Rhodococcus*." Environ Int 35(1): 162-177.

Mathieu, J. M., W. W. Mohn, L. D. Eltis, J. C. LeBlanc, G. R. Stewart, C. Dresen, K. Okamoto and P. J. Alvarez (2010). "7-ketocholesterol catabolism by *Rhodococcus jostii* RHA1." Appl Environ Microbiol 76(1): 352-355.

Mathieu, J. M., W. W. Mohn, L. D. Eltis, J. C. LeBlanc, G. R. Stewart, C. Dresen, K. Okamoto and P. J. Alvarez (2010). "7-ketocholesterol catabolism by *Rhodococcus jostii* RHA1." Appl. Environ. Microbiol. 76(1): 352-355.

Mathieu, J., J. Schloendorn, B. E. Rittmann and P. J. Alvarez (2008). "Microbial degradation of 7-ketocholesterol." Biodegradation 19(6): 807-813.

Matsui, T., H. Saeki, N. Shinzato and H. Matsuda (2006). "Characterization of *Rhodococcus-E. coli* shuttle vector pNC9501 constructed from the cryptic plasmid of a propene-degrading bacterium." Curr Microbiol 52(6): 445-448.

McLean, K. J., M. Hans, *et al.* (2012). "Cholesterol, an essential molecule: diverse roles involving cytochrome P450 enzymes." Biochem Soc Trans 40(3): 587-593.

McLeod, M. P., R. L. Warren, W. W. Hsiao, N. Araki, M. Myhre, C. Fernandes, D. Miyazawa, W. Wong, A. L. Lillquist, D. Wang, M. Dosanjh, H. Hara, A. Petrescu, R. D. Morin, G. Yang, J. M. Stott, J. E. Schein, H. Shin, D. Smailus, A. S. Siddiqui, M. A. Marra, S. J. Jones, R. Holt, F. S. Brinkman, K. Miyauchi, M. Fukuda, J. E. Davies, W. W. Mohn and L. D. Eltis (2006). "The complete genome of *Rhodococcus sp*. RHA1 provides insights into a catabolic powerhouse." Proc Natl Acad Sci USA 103(42): 15582-15587.

Mirzakhani, E. and F. M. Nejad (2016). "Grasses and *Rhodococcus erythropolis* Bacteria for Bioremediation of Naturally Polluted Soils with Hydrocarbons." Chemical Engineering & Technology 39(9): 1731-1737.

Mohn, W. W., M. H. Wilbrink, I. Casabon, G. R. Stewart, J. Liu, R. van der Geize and L. D. Eltis (2012). "Gene cluster encoding cholate catabolism in *Rhodococcus spp*." J Bacteriol 194(24): 6712-6719.

Mohn, W. W., R. van der Geize, G. R. Stewart, S. Okamoto, J. Liu, L. Dijkhuizen and L. D. Eltis (2008). "The actinobacterial mce4 locus encodes a steroid transporter." J. Biol. Chem. 283(51): 35368-35374.

Molnár, I., K. P. Choi, M. Yamashita and Y. Murooka (1995). "Molecular cloning, expression in Streptomyces lividans, and analysis of a gene cluster from *Arthrobacter simplex* encoding 3-ketosteroid-delta 1-dehydrogenase, 3-ketosteroid-delta 5-isomerase and a hypothetical regulatory protein." Mol. Microbiol. 15(5): 895-905.

Morii, S., C. Fujii, T. Miyoshi, M. Iwami and E. Itagaki (1998). "3-Ketosteroid-delta1-dehydrogenase of *Rhodococcus rhodochrous*: sequencing of the genomic DNA and hyperexpression, purification, and characterization of the recombinant enzyme." J Biochem 124(5): 1026-1032.

Murohisa, T. and M. Iida (1993). "Microbial degradation of 19-hydroxy-sterol side chains." Journal of fermentation and bioengineering 75(1): 13-17.

Murohisa, T. and M. Iida (1993). "Some new intermediates in microbial side chain degradation of β-sitosterol." Journal of fermentation and bioengineering 76(3): 174-177.

Nakano, C., A. Motegi, T. Sato, M. Onodera and T. Hoshino (2007). "Sterol biosynthesis by a prokaryote: first in vitro identification of the genes encoding squalene epoxidase and lanosterol synthase from *Methylococcus capsulatus*." Bioscience, biotechnology, and biochemistry 71(10): 2543-2550.

Nakashima, N. and T. Tamura (2004). "Isolation and characterization of a rolling-circle-type plasmid from *Rhodococcus erythropolis* and application of the plasmid to multiple-recombinant-protein expression." Appl Environ Microbiol 70(9): 5557-5568.

Navas, J., B. González-Zorn, N. Ladrón, P. Garrido and J. A. Vázquez-Boland (2001). "Identification and mutagenesis by allelic exchange of choE, encoding a cholesterol oxidase from the intracellular pathogen *Rhodococcus equi*." Journal of bacteriology 183(16): 4796-4805.

Nesbitt, N. M., X. Yang, P. Fontan, I. Kolesnikova, I. Smith, N. S. Sampson and E. Dubnau (2010). "A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstadienedione from cholesterol." Infect Immun 78(1): 275-282.

O'Brien, X., J. Parker, P. Lessard and A. Sinskey (2002). "Engineering an indene bioconversion process for the production of cis-aminoindanol: a model system for the production of chiral synthons." Applied microbiology and biotechnology 59(4): 389-399.

Ohyama, K., M. Suzuki, J. Kikuchi, K. Saito and T. Muranaka (2009). "Dual biosynthetic pathways to phytosterol via cycloartenol and lanosterol in *Arabidopsis*." Proceedings of the National Academy of Sciences 106(3): 725-730.

Ouellet, H., J. B. Johnston and P. R. de Montellano (2011). "Cholesterol catabolism as a therapeutic target in *Mycobacterium tuberculosis*." Trends Microbiol 19(11): 530-539.

Ouellet, H., S. Guan, *et al.* (2010). "*Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one." Mol Microbiol 77(3): 730-742.

Pandey, A. K. and C. M. Sassetti (2008). "Mycobacterial persistence requires the utilization of host cholesterol." Proc Natl Acad Sci U S A 105(11): 4376-4380.

Papadopoulos, J. S. and R. Agarwala (2007). "COBALT: constraint-based alignment tool for multiple protein sequences." Bioinformatics 23(9): 1073-1079.

Parales, R. E., J. V. Parales, *et al.* (1999). "Aspartate 205 in the catalytic domain of naphthalene dioxygenase is essential for activity." J Bacteriol 181(6): 1831-1837.

Patrauchan, M. A., C. Florizone, M. Dosanjh, W. W. Mohn, J. Davies and L. D. Eltis (2005). "Catabolism of benzoate and phthalate in *Rhodococcus sp.* strain RHA1: redundancies and convergence." J Bacteriol 187(12): 4050-4063.

Pearson, A., M. Budin and J. J. Brocks (2003). "Phylogenetic and biochemical evidence for sterol synthesis in the bacterium Gemmata obscuriglobus." Proceedings of the National Academy of Sciences 100(26): 15352-15357.

Penfield, J. S., L. J. Worrall, N. C. Strynadka and L. D. Eltis (2014). "Substrate specificities and conformational flexibility of 3-ketosteroid 9α-hydroxylases." Journal of Biological Chemistry 289(37): 25523-25536.

Perkins, C., S. Siddiqui, M. Puri and A. L. Demain (2016). "Biotechnological applications of microbial bioconversions." Critical reviews in biotechnology 36(6): 1050-1065.

Petersen, T. N., S. Brunak, G. von Heijne and H. Nielsen (2011). "SignalP 4.0: discriminating signal peptides from transmembrane regions." Nature Methods 8: 785-786.

Petrusma, M., G. Hessels, L. Dijkhuizen and R. van der Geize (2011). "Multiplicity of 3-ketosteroid-9alpha-hydroxylase enzymes in *Rhodococcus rhodochrous* DSM43269 for specific degradation of different classes of steroids." J Bacteriol 193(15): 3931-3940.

Petrusma, M., L. Dijkhuizen and R. van der Geize (2009). "*Rhodococcus rhodochrous* DSM 43269 3-ketosteroid 9alpha-hydroxylase, a two-component iron-sulfur-containing monooxygenase with subtle steroid substrate specificity." Appl Environ Microbiol 75(16): 5300-5307.

Petrusma, M., L. Dijkhuizen and R. van der Geize (2012). "Structural features in the KshA terminal oxygenase protein that determine substrate preference of 3-ketosteroid 9alpha-hydroxylase enzymes." J Bacteriol 194(1): 115-121.

Petrusma, M., R. van der Geize and L. Dijkhuizen (2014). "3-Ketosteroid 9α-hydroxylase enzymes: Rieske non-heme monooxygenases essential for bacterial steroid degradation." Antonie van Leeuwenhoek 106(1): 157-172.

Pfleiderer, C., A. Smid, I. Bartsch and I. Grummt (1990). "An undecamer DNA sequence directs termination of human ribosomal gene transcription." Nucleic Acids Res 18(16): 4727-4736.

Philipp, B. (2011). "Bacterial degradation of bile salts." Appl. Microbiol. Biotechnol. 89(4): 903-915.

Pieper, D. H. and M. Seeger (2008). "Bacterial metabolism of polychlorinated biphenyls." Journal of molecular microbiology and biotechnology 15(2-3): 121-138.

Pinto, A., M. Tarasev, *et al.* (2006). "Substitutions of the "bridging" aspartate 178 result in profound changes in the reactivity of the Rieske center of phthalate dioxygenase." Biochemistry 45(30): 9032-9041.

Plesiat, P., M. Grandguillot, S. Harayama, S. Vragar and Y. Michel-Briand (1991). "Cloning, sequencing, and expression of the *Pseudomonas testosteroni* gene encoding 3-oxosteroid delta 1-dehydrogenase." J Bacteriol 173(22): 7219-7227.

Płociniczak, T., E. Fic, M. Pacwa-Płociniczak, M. Pawlik and Z. Piotrowska-Seget (2017). "Improvement of phytoremediation of an aged petroleum hydrocarbon-contaminated soil by *Rhodococcus erythropolis* CD 106 strain." International Journal of Phytoremediation(just-accepted): 00-00.

Pollegioni, L., L. Piubelli and G. Molla (2009). "Cholesterol oxidase: biotechnological applications." FEBS J. 276(23): 6857-6870.

Poole, A. J. (2004). "Treatment of biorefractory organic compounds in wool scour effluent by hydroxyl radical oxidation." Water Res 38(14-15): 3458-3464.

Priefert, H., J. Rabenhorst and A. Steinbuchel (1997). "Molecular characterization of genes of *Pseudomonas sp.* strain HR199 involved in bioconversion of vanillin to protocatechuate." J Bacteriol 179(8): 2595-2607.

Priefert, H., X. M. O'Brien, P. A. Lessard, A. F. Dexter, E. E. Choi, S. Tomic, G. Nagpal, J. J. Cho, M. Agosto and L. Yang (2004). "Indene bioconversion by a toluene inducible dioxygenase of Rhodococcus sp. I24." Applied microbiology and biotechnology 65(2): 168-176.

Providenti, M. A., J. M. O'Brien, J. Ruff, A. M. Cook and I. B. Lambert (2006). "Metabolism of isovanillate, vanillate, and veratrate by *Comamonas testosteroni* strain BR6020." J Bacteriol 188(11): 3862-3869.

Reese, M. G. (2001). "Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome." Comput Chem 26(1): 51-56.

Rodríguez-García, A., E. Fernández-Alegre, *et al.* (2016). "Complete genome sequence of *'Mycobacterium neoaurum*' NRRL B-3805, an androstenedione (AD) producer for industrial biotransformation of sterols." J Biotechnol 224: 64-65.

Rohman, A., N. van Oosterwijk and B. W. Dijkstra (2012). "Purification, crystallization and preliminary X-ray crystallographic analysis of 3-ketosteroid Δ1-dehydrogenase from *Rhodococcus erythropolis* SQ1." Acta Crystallographica Section F: Structural Biology and Crystallization Communications 68(5): 551-556.

Rohman, A., N. van Oosterwijk, A.-M. W. Thunnissen and B. W. Dijkstra (2013). "Crystal structure and site-directed mutagenesis of 3-ketosteroid Δ1-dehydrogenase from *Rhodococcus erythropolis* SQ1 explain its catalytic mechanism." Journal of Biological Chemistry 288(49): 35559-35568.

Rosloniec, K. Z., M. H. Wilbrink, J. K. Capyk, W. W. Mohn, M. Ostendorf, R. van der Geize, L. Dijkhuizen and L. D. Eltis (2009). "Cytochrome P450 125 (CYP125) catalyses C26-hydroxylation to initiate sterol side-chain degradation in *Rhodococcus jostii* RHA1." Mol Microbiol 74(5): 1031-1043.

Ruprecht, A., J. Maddox, A. J. Stirling, N. Visaggio and S. Y. Seah (2015). "Characterization of novel acyl coenzyme A dehydrogenases involved in bacterial steroid degradation." Journal of bacteriology 197(8): 1360-1367.

Sallam, K. I., N. Tamura and T. Tamura (2007). "A multipurpose transposon-based vector system mediates protein expression in *Rhodococcus erythropolis*." Gene 386(1-2): 173-182.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular cloning: A Laboratory Manual. N.Y., Cold Spring Harbor Laboratory Press.

Schafer, A., A. Tauch, W. Jager, J. Kalinowski, G. Thierbach and A. Puhler (1994). "Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*." Gene 145(1): 69-73.

Schmid, R. D. and V. Urlacher (2007). Modern biooxidation: enzymes, reactions and applications, John Wiley & Sons.

Schneider, K., E. Graf, E. Irran, G. Nicholson, F. M. Stainsby, M. Goodfellow, S. A. Borden, S. Keller, R. D. Süssmuth and H.-P. Fiedler (2008). "Bendigoles A~ C, New Steroids from *Gordonia australis* Acta 2299 ." Journal of Antibiotics 61(6): 356.

Seth-Smith, H. M., J. Edwards, S. J. Rosser, D. A. Rathbone and N. C. Bruce (2008). "The explosive-degrading cytochrome P450 system is highly conserved among strains of *Rhodococcus spp.*" Applied and environmental microbiology 74(14): 4550-4552.

Shao, M., X. Zhang, Z. Rao, M. Xu, T. Yang, H. Li, Z. Xu and S. Yang (2016). "A mutant form of 3-ketosteroid-Delta(1)-dehydrogenase gives altered androst-1,4-diene-3, 17-dione/androst-4-ene-3,17-dione molar ratios in steroid biotransformations by Mycobacterium neoaurum ST-095." J Ind Microbiol Biotechnol 43(5): 691-701.

Sharma, M. (2014). "Actinomycetes: source, identification, and their applications." Int J Curr Microbiol Appl Sci 3(2): 801-832.

Shell, S. S., J. Wang, P. Lapierre, M. Mir, M. R. Chase, M. M. Pyle, R. Gawande, R. Ahmad, D. A. Sarracino, T. R. Ioerger, S. M. Fortune, K. M. Derbyshire, J. T. Wade and T. A. Gray (2015). "Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape." PLoS Genet 11(11): e1005641.

Shimizu, S., H. Kobayashi, E. Masai and M. Fukuda (2001). "Characterization of the 450-kb linear plasmid in a polychlorinated biphenyl degrader, *Rhodococcus sp.* strain RHA1." Appl Environ Microbiol 67(5): 2021-2028.

Shtratnikova, V. Y., M. I. Schelkunov, V. V. Fokina, Y. A. Pekov, T. Ivashina and M. V. Donova (2016). "Genome-wide bioinformatics analysis of steroid metabolism-associated genes in *Nocardioides simplex* VKM Ac-2033D." Curr Genet 62: 643-656.

Sigel, A., H. Sigel and R. K. Sigel (2007). The ubiquitous roles of cytochrome P450 proteins: metal ions in life sciences, John Wiley & Sons.

Silva-Rocha, R., E. Martinez-Garcia, B. Calles, M. Chavarria, A. Arce-Rodriguez, A. de Las Heras, A. D. Paez-Espino, G. Durante-Rodriguez, J. Kim, P. I. Nikel, R. Platero and V. de Lorenzo (2013). "The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes." Nucleic Acids Res 41(Database issue): D666-675.

Simon, R., U. Priefer and A. Pühler (1983). "A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria." Biotechnology 1: 748-791.

Singer, M. E. and W. R. Finnerty (1988). "Construction of an *Escherichia coli-Rhodococcus* shuttle vector and plasmid transformation in *Rhodococcus spp*." J Bacteriol 170(2): 638-645.

Söhngen, N. L. (1913). "Benzin, Petroleum, Parafinöll, und Paraffin als Kohlenstoff und Energiequelle für Microben." Zbl Bakteriol Parasitenk Abt II 37: 595-609.

Sokolovská, I., R. Rozenberg, C. Riez, P. G. Rouxhet, S. N. Agathos and P. Wattiau (2003). "Carbon source-induced modifications in the mycolic acid content and cell wall permeability of *Rhodococcus erythropolis* E1." Applied and environmental microbiology 69(12): 7019-7027.

Solovyev, V. and A. Salamov, Eds. (2011). Automatic Annotation of Microbial Genomes and Metagenomic Sequences. Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies, Nova Science Publishers.

Soto, A. M., J. M. Calabro, N. V. Prechtl, A. Y. Yau, E. F. Orlando, A. Daxenberger, A. S. Kolok, L. J. Guillette, Jr., B. le Bizec, I. G. Lange and C. Sonnenschein (2004). "Androgenic and estrogenic activity in water bodies receiving cattle feedlot effluent in Eastern Nebraska, USA." Environ Health Perspect 112(3): 346-352.

Strohl, W. R. (1992). "Compilation and analysis of DNA sequences associated with apparent streptomycete promoters." Nucleic Acids Res 20(5): 961-974.

Suemori, A., K. Nakajima, R. Kurane and Y. Nakamura (1995). "o-, m- and p-hydroxybenzoate degradative pathways in *Rhodococcus erythropolis*." FEMS Microbiol Lett 125(1): 31-35.

Sukhodolskaya, G. V., V. M. Nikolayeva, S. M. Khomutov and M. V. Donova (2007). "Steroid-1-dehydrogenase of *Mycobacterium sp.* VKM Ac-1817D strain producing 9alpha-hydroxy-androst-4-ene-3,17-dione from sitosterol." Appl Microbiol Biotechnol 74(4): 867-873.

Sumpter, J. P. and S. Jobling (1995). "Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment." Environmental health perspectives 103(Suppl 7): 173.

Sutcliffe, I. C., A. K. Brown and L. G. Dover (2010). The Rhodococcal Cell Envelope: Composition, Organisation and Biosynthesis. Berlin Heidelberg, Springer.

Swain, K., I. Casabon, L. D. Eltis and W. W. Mohn (2012). "Two transporters essential for reassimilation of novel cholate metabolites by *Rhodococcus jostii* RHA1." J. Bacteriol. 194(24): 6720-6727.

Szoköl, J., L. Rucká, M. Simciková, P. Halada, J. Nesvera and M. Pátek (2014). "Induction and carbon catabolite repression of phenol degradation genes in *Rhodococcus erythropolis* and *Rhodococcus jostii*." Appl Microbiol Biotechnol 98(19): 8267-8279.

Tabas, I. (2002). "Cholesterol in health and disease." J Clin Invest 110: 583-590.

Takeo, M., M. Murakami, S. Niihara, K. Yamamoto, M. Nishimura, D.-i. Kato and S. Negoro (2008). "Mechanism of 4-nitrophenol oxidation in *Rhodococcus sp.* Strain PN1: characterization of the two-component 4-nitrophenol hydroxylase and regulation of its expression." Journal of bacteriology 190(22): 7367-7374.

Tarasev, M. and D. P. Ballou (2005). "Chemistry of the catalytic conversion of phthalate into its cis-dihydrodiol during the reaction of oxygen with the reduced form of phthalate dioxygenase." Biochemistry 44(16): 6197-6207.

Tarasev, M., A. Pinto, *et al.* (2006). "The "bridging" aspartate 178 in phthalate dioxygenase facilitates interactions between the Rieske center and the iron (II)--mononuclear center." Biochemistry 45(34): 10208-10216.

Thomas, S. T. and N. S. Sampson (2013). "*Mycobacterium tuberculosis* utilizes a unique heterotetrameric structure for dehydrogenation of the cholesterol side chain." Biochemistry 52(17): 2895-2904.

Thomas, S. T., B. C. Vanderven, D. R. Sherman, D. G. Russell and N. S. Sampson (2011). "Pathway profiling in *Mycobacterium tuberculosis*: Elucidation of cholesterol-derived catabolite and enzymes that catalyze its metabolism." J Biol Chem 286(51): 43668-43678.

Tomas-Gallardo, L., H. Gomez-Alvarez, E. Santero and B. Floriano (2014). "Combination of degradation pathways for naphthalene utilization in *Rhodococcus sp.* strain TFB." Microb Biotechnol 7(2): 100-113.

Tong, W.-Y. and X. Dong (2009). "Microbial biotransformation: recent developments on steroid drugs." Recent patents on biotechnology 3(2): 141-153.

Travkin, V. M., I. P. Solyanikova and L. A. Golovleva (2006). "Hydroxyquinol pathway for microbial degradation of halogenated aromatic compounds." J Environ Sci Health B 41(8): 1361-1382.

Uhía, I., B. Galán, F. J. Medrano and J. L. García (2011). "Characterization of the KstR-dependent promoter of the gene for the first step of the cholesterol degradative pathway in *Mycobacterium smegmatis*." Microbiology 157(Pt 9): 2670-2680.

Uhía, I., B. Galán, S. L. Kendall, N. G. Stoker and J. L. García (2012). "Cholesterol metabolism in *Mycobacterium smegmatis*." Environ Microbiol Reports 4(2): 168-182.

Uhía, I., B. Galán, V. Morales and J. García (2011). "Initial step in the catabolism of cholesterol by *Mycobacterium smegmatis* mc2155." Environmental microbiology 13(4): 943-959.

Uhía, I., B. Galan, V. Morales and J. L. García (2011). "Initial step in the catabolism of cholesterol by *Mycobacterium smegmatis* mc2 155." Environ Microbiol 13(4): 943-959.

Urai, M., H. Anzai, J. Ogihara, N. Iwabuchi, S. Harayama, M. Sunairi and M. Nakajima (2006). "Structural analysis of an extracellular polysaccharide produced by *Rhodococcus rhodochrous* strain S-2." Carbohydrate research 341(6): 766-775.

Urai, M., H. Yoshizaki, H. Anzai, J. Ogihara, N. Iwabuchi, S. Harayama, M. Sunairi and M. Nakajima (2007). "Structural analysis of mucoidan, an acidic extracellular polysaccharide produced by a pristane-assimilating marine bacterium, *Rhodococcus erythropolis* PR4." Carbohydrate research 342(7): 927-932.

Valderrama, J. A., G. Durante-Rodriguez, B. Blazquez, J. L. Garcia, M. Carmona and E. Diaz (2012). "Bacterial degradation of benzoate: cross-regulation between aerobic and anaerobic pathways." J Biol Chem 287(13): 10494-10508.

van der Geize, R. and L. Dijkhuizen (2004). "Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications." Curr Opin Microbiol 7(3): 255-261.

van der Geize, R., A. Grommen, G. Hessels, A. Jacobs and L. Dijkhuizen (2011). "The steroid catabolic pathway of the intracellular pathogen *Rhodococcus equi* is important for pathogenesis and a target for vaccine development." PLoS Pathog 7(8): e1002181.

van Der Geize, R., G. Hessels and L. Dijkhuizen (2008). Method for the production of modified steroid degrading microorganisms and their use, Google Patents.

van der Geize, R., G. I. Hessels and L. Dijkhuizen (2002). "Molecular and functional characterization of the kstD2 gene of *Rhodococcus erythropolis* SQ1 encoding a second 3-ketosteroid Delta(1)-dehydrogenase isoenzyme." Microbiology 148(Pt 10): 3285-3292.

van der Geize, R., G. I. Hessels, M. Nienhuis-Kuiper and L. Dijkhuizen (2008). "Characterization of a second *Rhodococcus erythropolis* SQ1 3-ketosteroid 9alpha-hydroxylase activity comprising a terminal oxygenase homologue, KshA2, active with oxygenase-reductase component KshB." Appl Environ Microbiol 74(23): 7197-7203.

van der Geize, R., G. I. Hessels, R. van Gerwen, J. W. Vrijbloed, P. van der Meijden and L. Dijkhuizen (2000). "Targeted disruption of the kstD gene encoding a 3-ketosteroid-Δ1-dehydrogenase isoenzyme of *Rhodococcus erythropolis* strain SQ1." Appl Environ Microbiol 66(5): 2029-2036.

van der Geize, R., G. I. Hessels, R. van Gerwen, P. van der Meijden and L. Dijkhuizen (2001). "Unmarked gene deletion mutagenesis of kstD, encoding 3-ketosteroid Delta1-dehydrogenase, in *Rhodococcus erythropolis* SQ1 using sacB as counter-selectable marker." FEMS Microbiol Lett 205(2): 197-202.

van der Geize, R., G. I. Hessels, R. van Gerwen, P. van der Meijden and L. Dijkhuizen (2002). "Molecular and functional characterization of kshA and kshB, encoding two components of 3-ketosteroid 9alpha-hydroxylase, a class IA monooxygenase, in *Rhodococcus erythropolis* strain SQ1." Mol Microbiol 45(4): 1007-1018.

van der Geize, R., K. Yam, T. Heuser, M. H. Wilbrink, H. Hara, M. C. Anderton, E. Sim, L. Dijkhuizen, J. E. Davies, W. W. Mohn and L. D. Eltis (2007). "A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages." Proc. Natl. Acad. Sci. U.S.A. 104(6): 1947-1952.

van Oosterwijk, N., J. Knol, L. Dijkhuizen, R. van der Geize and B. W. Dijkstra (2011). "Cloning, overexpression, purification, crystallization and preliminary X-ray analysis of 3-ketosteroid Δ4-(5α)-dehydrogenase from *Rhodococcus jostii* RHA1." Acta Crystallographica Section F: Structural Biology and Crystallization Communications 67(10): 1269-1273.

van Oosterwijk, N., J. Knol, L. Dijkhuizen, R. van der Geize and B. W. Dijkstra (2012). "Structure and catalytic mechanism of 3-ketosteroid-Δ4-(5α)-dehydrogenase from *Rhodococcus jostii* RHA1 genome." Journal of Biological Chemistry 287(37): 30975-30983.

Vásquez, T. G. P., A. E. C. Botero, L. M. S. De Mesquita and M. L. Torem (2007). "Biosorptive removal of Cd and Zn from liquid streams with a *Rhodococcus opacus* strain." Minerals engineering 20(9): 939-944.

Veeranagouda, Y., E. J. Lim, D. W. Kim, J.-K. Kim, K. Cho, H. J. Heipieper and K. Lee (2009). "Formation of specialized aerial architectures by *Rhodococcus* during utilization of vaporized p-cresol." Microbiology 155(11): 3788-3796.

Veiga, P., C. Juste, P. Lepercq, K. Saunier, F. Beguet and P. Gerard (2005). "Correlation between faecal microbial community structure and cholesterol-to-coprostanol conversion in the human gut." FEMS Microbiol Lett 242(1): 81-86.

Venkataraman, H., E. M. Te Poele, *et al.* (2015). "Biosynthesis of a steroid metabolite by an engineered *Rhodococcus erythropolis* strain expressing a mutant cytochrome P450 BM3 enzyme." Appl Microbiol Biotechnol 99(11): 4713-4721.

Ventura, M., C. Canchaya, A. Tauch, G. Chandra, G. F. Fitzgerald, K. F. Chater and D. van Sinderen (2007). "Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum." Microbiology and Molecular Biology Reviews 71(3): 495-548.

Voges, D., P. Zwickl and W. Baumeister (1999). "The 26S proteasome: a molecular machine designed for controlled proteolysis." Annu Rev Biochem 68: 1015-1068.

Von Bargen, K. and A. Haas (2009). "Molecular and infection biology of the horse pathogen *Rhodococcus equi*." FEMS microbiology reviews 33(5): 870-891.

Vrielink, A. and S. Ghisla (2009). "Cholesterol oxidase: biochemistry and structural features." FEBS J. 276(23): 6826-6843.

Wagner, B., P. G. Atrat, J. E. Clark-Curtiss and M. Wagner (1992). "Localization of the steroid 1-dehydrogenase in *Rhodococcus erythropolis* IMET 7030 by immunoelectron microscopy." J Basic Microbiol 32(1): 65-71.

Wang, C., J. Lee, Y. Deng, F. Tao and L. H. Zhang (2012). "ARF-TSS: an alternative method for identification of transcription start site in bacteria." Biotechniques 52(4).

Wang, F. Q., K. Yao, *et al.* (2011). Agricultural and Biological Sciences "Soybean and Health". Chapter 11: From soybean phytosterols to steroid hormones. Soybean Health, InTech, Open access Publisher, Rijeka, Hany El-Shemy (Ed.), ISBN 978-953-307-535-8.

Wang, P. H., T. H. Lee, *et al.* (2013). "An oxygenase-independent cholesterol catabolic pathway operates under oxic conditions." PLoS One 8(6): e66675. doi: 10.1371/journal.pone.0066675.

Warren, R., W. W. Hsiao, H. Kudo, M. Myhre, M. Dosanjh, A. Petrescu, H. Kobayashi, S. Shimizu, K. Miyauchi and E. Masai (2004). "Functional characterization of a catabolic plasmid from polychlorinated-biphenyl-degrading *Rhodococcus sp.* strain RHA1." Journal of bacteriology 186(22): 7783-7795.

Watanabe, K., H. Shimizu, H. Aihara, R. Nakamura, K.-i. Suzuki and K. Komagata (1986). "Isolation and identification of cholesterol-degrading *Rhodococcus* strains from food of animal origin and their cholesterol oxidase activities." The Journal of General and Applied Microbiology 32(2): 137-147.

Wei, W., F. Q. Wang, S. Y. Fan and D. Z. Wei (2010). "Inactivation and augmentation of the primary 3-ketosteroid-{delta}1- dehydrogenase in *Mycobacterium neoaurum* NwIB-01: biotransformation of soybean phytosterols to 4-androstene- 3,17-dione or 1,4-androstadiene-3,17-dione." Appl Environ Microbiol 76(13): 4578-4582.

Wei, W., S.-Y. Fan, F.-Q. Wang and D.-Z. Wei (2014). "Accumulation of androstadiene-dione by overexpression of heterologous 3-ketosteroid Δ1-dehydrogenase in *Mycobacterium neoaurum* NwIB-01." World Journal of Microbiology and Biotechnology 30(7): 1947-1954.

Wierenga, R. K., P. Terpstra and W. G. Hol (1986). "Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint." J Mol Biol 187(1): 101-107.

Wilbrink, M. H., M. Petrusma, *et al.* (2011). "FadD19 of *Rhodococcus rhodochrous* DSM43269, a steroid-coenzyme A ligase essential for degradation of C-24 branched sterol side chains." Appl Environ Microbiol 77(13): 4455-4464.

Wrońska, N., A. Brzostek, R. Szewczyk, A. Soboń, J. Dziadek and K. Lisowska (2016). "The Role of fadD19 and echA19 in Sterol Side Chain Degradation by *Mycobacterium smegmatis*." Molecules 21(5): 598.

Xu, L.-Q., Y.-J. Liu, K. Yao, H.-H. Liu, X.-Y. Tao, F.-Q. Wang and D.-Z. Wei (2016). "Unraveling and engineering the production of 23, 24-bisnorcholenic steroids in sterol metabolism." Scientific reports 6.

Xu, X. W., X. Q. Gao, J. X. Feng, X. D. Wang and D. Z. Wei (2015). "Influence of temperature on nucleus degradation of 4-androstene-3, 17-dione in phytosterol biotransformation by *Mycobacterium sp*." Lett Appl Microbiol 61(1): 63-68.

Yachdav, G., E. Kloppmann, L. Kajan, M. Hecht, T. Goldberg, T. Hamp, P. Honigschmid, A. Schafferhans, M. Roos, M. Bernhofer, L. Richter, H. Ashkenazy, M. Punta, A. Schlessinger, Y. Bromberg, R. Schneider, G. Vriend, C. Sander, N. Ben-Tal and B. Rost (2014). "PredictProtein--an open resource for online prediction of protein structural and functional features." Nucleic Acids Res 42(Web Server issue): W337-343.

Yang, J. C., P. A. Lessard and A. J. Sinskey (2007). "Characterization of the mobilization determinants of pAN12, a small replicon from *Rhodococcus erythropolis* AN12." Plasmid 57(1): 71-81.

Yam, K. C. and S. Okamoto (2011). "Adventures in *Rhodococcus* - from steroids to explosives." Can J Microbiol 57(3): 155-168.

Yam, K. C., I. D'Angelo, R. Kalscheuer, H. Zhu, J. X. Wang, V. Snieckus, L. H. Ly, P. J. Converse, W. R. Jacobs, N. Strynadka and L. D. Eltis (2009). "Studies of a ring-cleaving dioxygenase illuminate the role of cholesterol metabolism in the pathogenesis of *Mycobacterium tuberculosis.*" PLoS Pathog 5(3): e1000344.

Yam, K. C., R. van der Geize and L. D. Eltis (2010). "Catabolism of Aromatic Compounds and Steroids by *Rhodococcus*". Berlin Heidelberg, Springer-Verlag.

Yang, J. C., P. A. Lessard, N. Sengupta, S. D. Windsor, X. M. O'Brien, M. Bramucci, J.-F. Tomb, V. Nagarajan and A. J. Sinskey (2007). "TraA is required for megaplasmid conjugation in *Rhodococcus erythropolis* AN12." Plasmid 57(1): 55-70.

Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang (2015). "The I-TASSER Suite: protein structure and function prediction." Nat Methods 12(1): 7-8.

Yang, M., K. E. Guja, S. T. Thomas, M. Garcia-Diaz and N. S. Sampson (2014). "A Distinct MaoC-like enoyl-CoA hydratase architecture mediates cholesterol catabolism in *Mycobacterium tuberculosis.*" ACS chemical biology 9(11): 2632-2645.

Yang, M., R. Lu, K. E. Guja, M. F. Wipperman, J. R. St. Clair, A. C. Bonds, M. Garcia-Diaz and N. S. Sampson (2015). "Unraveling cholesterol catabolism in *Mycobacterium tuberculosis*: ChsE4-ChsE5 α2β2 acyl-CoA dehydrogenase initiates β-oxidation of 3-oxo-cholest-4-en-26-oyl CoA." ACS infectious diseases 1(2): 110-125.

Yang, X., E. Dubnau, I. Smith and N. S. Sampson (2007). "Rv1106c from *Mycobacterium tuberculosis* is a 3beta-hydroxysteroid dehydrogenase." Biochemistry 46(31): 9058-9067.

Yang, X., J. Gao, I. Smith, E. Dubnau and N. S. Sampson (2011). "Cholesterol is not an essential source of nutrition for *Mycobacterium tuberculosis* during infection." Journal of bacteriology 193(6): 1473-1476.

Yang, X., N. M. Nesbitt, E. Dubnau, I. Smith and N. S. Sampson (2009). "Cholesterol metabolism increases the metabolic pool of propionate in Mycobacterium tuberculosis." Biochemistry 48(18): 3819-3821.

Yao, K., L.-Q. Xu, F.-Q. Wang and D.-Z. Wei (2014). "Characterization and engineering of 3-ketosteroid-Δ1-dehydrogenase and 3-ketosteroid-9α-hydroxylase in *Mycobacterium neoaurum* ATCC 25795 to produce 9α-hydroxy-4-androstene-3, 17-dione through the catabolism of sterols." Metabolic engineering 24: 181-191.

Yeh, C. H., Y. S. Kuo, C. M. Chang, W. H. Liu, M. L. Sheu and M. Meng (2014). "Deletion of the gene encoding the reductase component of 3-ketosteroid 9α-hydroxylase in *Rhodococcus equi* USA-18 disrupts sterol catabolism, leading to the accumulation of 3-oxo-23,24-bisnorchola-1,4-dien-22-oic acid and 1,4-androstadiene-3,17-dione." Microb Cell Fact 13: 130.

Yuan, J., G. Chen, *et al.* (2015). "Accumulation of 9alpha-hydroxy-4-androstene-3,17-dione by co-expressing kshA and kshB encoding component of 3-ketosteroid-9alpha-hydroxylase in *Mycobacterium sp.* NRRL B-3805." Sheng Wu Gong Cheng Xue Bao 31(4): 523-533.

Yue, Q. K., I. J. Kass, N. S. Sampson and A. Vrielink (1999). "Crystal structure determination of cholesterol oxidase from *Streptomyces* and structural characterization of key active site mutants." Biochemistry 38(14): 4277-4286.

Zhang, Q., Y. Ren, J. He, S. Cheng, J. Yuan, F. Ge, W. Li, Y. Zhang and G. Xie (2015). "Multiplicity of 3-ketosteroid Δ1-dehydrogenase enzymes in *Gordonia neofelifaecis* NRRL B-59395 with preferences for different steroids." Annals of microbiology 65(4): 1961-1971.

Zhang, X., D. Wu, T. Yang, M. Xu and Z. Rao (2016). "Over-expression of *Mycobacterium neoaurum* 3-ketosteroid-Δ1-dehydrogenase in *Corynebacterium crenatum* for efficient bioconversion of 4-androstene-3, 17-dione to androst-1, 4-diene-3, 17-dione." Electronic Journal of Biotechnology 24: 84-90.

Zhang, X., Z. Rao, L. Zhang, M. Xu and T. Yang (2016). "Efficient 9α-hydroxy-4-androstene-3, 17-dione production by engineered *Bacillus subtilis* co-expressing *Mycobacterium neoaurum* 3-ketosteroid 9α-hydroxylase and *B. subtilis* glucose 1-dehydrogenase with NADH regeneration." SpringerPlus 5(1): 1207.

Zhi, X.-Y., W.-J. Li and E. Stackebrandt (2009). "An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa." International Journal of Systematic and Evolutionary Microbiology 59(3): 589-608.

X. ANEXO

LISTADO DE PUBLICACIONES

Guevara, G., L. F. de las Heras, J. Perera and J. M. N. Llorens (2017). "Functional differentiation of 3-ketosteroid Δ 1-dehydrogenase isozymes in *Rhodococcus ruber* strain Chol-4." Microbial Cell Factories 16(1): 42.

Guevara, G., L. F. de las Heras, J. Perera and J. M. N. Llorens (2017). "Characterization of 3-Ketosteroid 9α-Hydroxylases in *Rhodococcus ruber* strain Chol-4". En revisión en The Journal of Steroid Biochemistry and Molecular Biology.

Guevara, G., J. Perera and J. M. N. Llorens (2016). Analysis of intermediates of steroid transformations in resting cells by thin layer chromatography (TLC). Capítulo de libro Methods in Molecular Biology tema MICROBIAL STEROIDS.